DOCUMENT RESUME

ED 038 873 EM 007 987

AUTHOR Glaser, Robert; Nitko, Anthony J.
TITLE Measurement in Learning and Instruction.
INSTITUTION Pittsburgh Univ., Pa. Learning Research and Development Center.
SPONS AGENCY Office of Naval Research, Washington, D.C. Psychological Sciences Div.
PUB DATE Mar 70
NOTE 129p.

EDRS PRICE EDRS Price MF-$0.75 HC-$6.55
DESCRIPTORS *Instructional Design, Learning Characteristics, *Measurement Techniques, Models, Program Evaluation, Test Construction

ABSTRACT

Measurement in learning and instruction is discussed here in the light of instructional design requirements and specific models or systems of instruction. Three general classes of instructional models found in current educational practice are presented. One particular model of instruction--for adapting instruction to individual differences--is described, and its testing and measurement implications are discussed. The description of the instructional model is followed by considerations of (a) the analysis of performance domains, (b) individual assignment to instructional alternatives, and (c) measuring what is learned by means of criterion-referenced tests. These topics are discussed in terms of the measurements required to make instructional decisions about individual learners. In the last section, the topic of evaluating and improving an instructional system and its components is discussed. A list of references, tables, and illustrations are appended. (Author/JY)

UNIVERSITY OF PITTSBURGH - LEARNING R & D CENTER

TECHNICAL REPORT

MEASUREMENT IN LEARNING AND INSTRUCTION

ROBERT GLASER AND ANTHONY J. NITKO

ED0 38873

MEASUREMENT IN LEARNING AND INSTRUCTION

Robert Glaser and Anthony J. Nitko

Learning Research and Development Center

University of Pittsburgh

March, 1970

The research reported herein was performed pursuant to Contract Nonr-624(18) with the Personnel and Training Branch, Psychological Sciences Division, Office of Naval Research. The document is a publication of the Learning Research and Development Center, supported in part as a research and development center by funds from the United States Office of Education, Department of Health, Education and Welfare.

Contents

MEASUREMENT IN LEARNING AND INSTRUCTION[1]

Robert Glaser and Anthony J. Nitko

University of Pittsburgh

With respect to the educational process, learning is defined as the acquisition of behavior brought about by the school environment and instructional means designed by the educator and the educational system. Ideally, the learner interacts with the instructional environment, changes it, and is changed in turn by the consequences of his actions. The particular properties of the behavior acquired by the learner depend upon the details of the educational environment that is designed and provided. What is taught and how it is taught depend upon the objectives and values of the school system; what and how, however, are not separable questions. The instructional environment can influence the student's behavior more or less directly: It can enable the student to acquire certain kinds of performance, and it can teach him to teach himself. Fostering, nurturing, guiding, influencing, and controlling human behavior is the practical objective of the educational enterprise. Educational environments designed and provided by society influence and control student behavior; they cannot do otherwise since the existence of any environment, whether it be a culture, a home, or a school, shapes behavior in intended and unintended ways. Many facets of human behavior are involved: the learning of subject-matter content and of the skills and processes involved in using it,

e.g., retention, transfer, problem-solving, critical thinking, creating, ways of processing information, and attitudes and motivation toward these activities. The design of an educational environment is a complex and subtle enterprise, and different kinds of environments encourage the occurrence of certain kinds of behavior and minimize and discourage others.

Testing and measurement are critical components of the educational environment--they provide the essential information for the development, operation, and evaluation of the educational enterprise. To be useful, this information must be relevant to the specific instructional system with which one is concerned. That is, information requirements are derived from and specified by an analysis of a particular educational environment, and are unique to it. Different educational environments will have different informational requirements. This is not to say, however, that a particular instructional system needs information only about itself. For example, the values and goals of other systems may inform the particular system at hand. It should be clear then, that since testing and measurement provide unique and relevant information, the design of testing and measurement procedures must be preceded by the specification of the particular instructional system (and the information requirements) for which these procedures are intended. What needs to be measured is then known, insofar as possible, and a testing program can be designed to satisfy these requirements. In short, measurement procedures need to be designed with the information requirements of a specific instructional system in mind.

The fundamental task of testing and measurement in education is to provide information for making decisions about instructional design and operation. Four activities are involved: analysis of the subject-matter domain under consideration, diagnosis of the characteristics of the learner, design of the instructional environment, and evaluation of learning outcomes.

In the analysis of the subject-matter domain, subject-matter experts are assisted in analyzing their domains in terms of the performance competencies which comprise them. Representative instances of competent performance are analyzed according to the properties of the content involved and the ways in which a student must respond to and process this content. The structural characteristics of the domain are laid out according to its conceptual hierarchies and operating rules in terms of increasing complexity of student performance. Major concerns are the analysis and definition of instructionally relevant performance, including the specification of educational objectives, translating these objectives into some kind of assessable performance, and performing studies and gathering data about the facilitating or inhibiting effects of particular curriculum sequences. The kind of analysis that goes on at this time is a significant determinant of the subsequent stages of instructional design. Learning is analyzed in terms of its subject-matter content and also in terms of the behavioral repertoires or behavioral processes that are being learned. These properties of content and process define the nature of measuring instruments and the nature of instruction.

The second activity, diagnosis of the characteristics of the learner, involves measurement of the behavior with which a student enters into instruction, including (a) the extent to which the student has already acquired what is to be learned, (b) the extent to which he has the necessary prerequisites, and (c) the characteristics of the way in which he learns that interact with the available instructional alternatives. These measurements provide information about the <u>existing pre-instructional behavior</u> of the learner as distinguished from the performance competence <u>to be achieved</u>. When attempting to provide this kind of information, one is concerned with the problems that arise in the measurement of individual differences. However, for instructional purposes, the concern reduces to those differences that are especially relevant to the instructional system that has been devised. No doubt, different individual capabilities require different modes of instruction. The general problem is the interaction between individual differences and the instructional environment [Ed: cross-reference to Cronbach's chapter, pp. 148 ff.]. It is increasingly apparent that for effective instruction, measurements must be made of differences in learning characteristics. The kinds of measurements that need to be taken will differ depending upon the options available in the instructional system. Characteristics that will predict the success of students in a relatively fixed environment will be different from those of students in a system where there are multiple paths to the same end.

Once the nature of the task to be learned and the entering characteristics of the learner are described, the third activity--designing

the instructional environment--can take place. The design of the instructional environment involves the specification and provision of the conditions under which learning can occur--that is, conditions that allow the learner to progress from an entering-behavior state to the terminal state of subject-matter competence. This activity includes the design and construction of teaching procedures, materials, and tests that are to be employed in the educational process. Also included are provisions for motivation to use, maintain, and extend the competence that is taught. The information required for the design and construction of the learning environment has two purposes. One is information for modifying decisions about how instruction is to proceed; the other is information for the design of instructional procedures, materials, and equipment. With regard to the first, as instruction proceeds, information for instructional decisions must be provided to the teacher, the student, and possibly to a machine, each of which assists in guiding the student through the course of instruction. In light of present educational innovation, it is highly likely that the job of the teacher will be influenced by procedures which allow assessment decisions to be made increasingly by the student himself and also by computer testing and related instructional devices. (The design of tests for use by the student in self-assessment has been seriously neglected in the past by educational test constructors.) With respect to the second kind of information, testing and measurement activities will also be required to support the adoption of innovative techniques and to support the maintenance of worthwhile, existing techniques. Just as, at the present

time, commercially available tests must present evidence about their development and documentation of their effectiveness, so will instructional techniques--whether they be procedures or devices--need to be accompanied by information to support their construction and improvement and to document their effectiveness.

Finally, the fourth activity--evaluating learning outcomes--involves assessing (a) the extent to which the acquired behavior of the learner approaches performance criteria, and (b) the extent to which the values espoused by the designers of the system and associated with this performance have been attained. Thus, the primary requirement is for measurement of what has been learned. The "what is learned" becomes fundamental since the instructional process requires information about the details of the performance of the learner in order to know how instruction should proceed. "What" includes both content and process and is defined, insofar as possible, with reference to prespecified performance criteria. When this performance has been attained by an individual learner to the degree required by the designers of the instructional system, then the learner is said to have attained _mastery_ of the instructional goal. Measurements that provide this kind of information may be termed absolute measurements [Ed.: cross-reference Cronbach's chapter pp. 11 ff.] and the tests constructed with this kind of measurement in mind are called criterion-referenced tests (the reader should refer to pp. 56 ff. of this chapter where a more formal definition of criterion-referenced tests is developed and where their construction requirements are discussed). Performance referenced only by

norms does not define what is learned; therefore, appropriate information is not provided about what individuals can do and how they behave. The information necessary for instructional decision-making is essentially descriptive of present performance (that is, at the time of testing) and is not predictive in the sence of predictive validity. The major predictive concern in the measurement of learning outcomes is the relationship between proximate and ultimate educational objectives, and this is more of a learning transfer problem than a correlational one.[2]

To recapitulate, learning in the educational sense can be defined as a process of transition of the learner from an initial entering state to a specified arbitrary terminal state. Instruction and teaching are the practices in schools by which conditions are provided to enable this transition to occur. Measurement in instruction and learning is concerned with providing data, assessments, and information about the nature of learner performance and about the nature of instructional conditions. The assessment of student performance is used to guide the implementation of appropriate instructional conditions, and the measurement of the conditions is used to indicate whether the conditions are, indeed, realized. In addition to guiding the instructional process, measurement is used to evaluate its total effectiveness. All these measurements are used for making decisions in the course of developing an instructional system, during its operation, and after it has occurred to evaluate its overall outcomes.

The Approach of This Chapter

As the above introductory comments suggest, measurement in learning and instruction should be discussed in light of instructional design requirements and specific models or systems of instruction. We approach this task as follows: Initially, three general classes of instructional models found in current educational practice are presented. One particular model of instruction--for adapting instruction to individual differences--is described, and its testing and measurement implications are discussed. The description of the instructional model is followed by considerations of (a) the analysis of performance domains, (b) individual assignment to instructional alternatives, and (c) measuring what is learned by means of criterion-referenced tests. These topics are discussed in terms of the measurements required to make instructional decisions about individual learners. In the last section, the important topic of evaluating and improving an instructional system and its components is discussed. At that point, group-learner data play a more central role.

The reader should note that throughout the first part of this chapter, measurements and tests which provide information relevant to absolute decision-making are called for. [Ed.: cross-reference to Cronbach's chapter, pp. 11 ff.] The design, construction, and use of such tests justify a more detailed treatment than that provided in the course of discussing the overall testing requirements of the particular instructional model examined. As a consequence, a separate section

(pp. 56 ff.) dealing with criterion-referenced tests is provided. It is hoped that the initial considerations of the measurements which are required in the context of an instructional system will serve as an "advanced organizer" on the subject of criterion-referenced testing. The reader who feels that his needs are best served by first examining the more detailed treatment of this type of test may read the later section first without loss of continuity.

To place the topics of this chapter into perspective, a brief review is presented of the way in which the relationships among the disciplines of psychological measurement, experimental psychology, and the field of educational practice have influenced the state of measurement in learning and instruction.

Some History

A significant complication in the field of measurement in learning and instruction results from the historical routes of two major fields of psychology: the measurement of individual differences and the experimental psychology of learning. It is well documented that early scientific psychology began with these two as apparently separate disciplines. This history can be traced from the Titchener-Baldwin controversy in the 1890's, through Cronbach's (1957) address on "The Two Disciplines of Scientific Psychology," through the 1967 book edited by Gagné on Learning and Individual Differences. Throughout the years, the importance of coordination between the two fields has been recognized, but with sustained work by only a few individuals. The require-

ments inherent in developing a scientific base for instruction make this coordination mandatory, with changes in traditional practices being required in each field. E. L. Thorndike (1914) raised the problem in his Educational Psychology pointing to experiments that showed the effect of equal learning opportunity, i.e., equal practice, on producing increases or decreases in individual differences. Woodrow (1938) pointed out that the divergence or convergence of individual differences with practice depended upon the shape of the learning curve and the position of individuals on it as a result of their prior task-relevant experience. In addition, Woodrow indicated that the influence of individual differences in the course of practice might also be a function of the way in which the task changes during practice. Recent work on this problem has been carried out in a series of studies by Fleishman (1967), which show that final scores in a learning task are dependent upon a different pattern of abilities than initial scores.

In a classic article, Woodrow (1946) pointed out the lack of relationship between general ability measures, such as intelligence, and learning variables. Woodrow's findings, from both the laboratory and the classroom, contradicted the assumption that the ability to learn, in the sense of ability to improve with practice, is related to measured intelligence. Correlations between intelligence and gain were generally not statistically significant. Woodrow interpreted his results by assuming that a score at any stage of practice consists of a general factor, G, and specific factors, these latter changing with practice. As a result, there can be a high and undiminishing correla-

tion between the general factor and scores at all stages of practice; it is also possible for the correlation between G and gain to be negligible when gain is the result of a high degree of specificity resulting from task characteristics and individual differences in performing these tasks. The line of work generated by Woodrow has been reflected in the active interest of this problem by DuBois (   ) and by Gulliksen and his students, e.g., Stake (1961) and Duncanson (1964).

On the side of learning theory, Hull (1945), in developing his theory of learning, initially gave serious attention to individual differences in learning. He pointed out that the study of behavior has two tasks: the first is deriving primary laws as displayed by the model or average organism under given conditions; the second is the problem of innate behavioral differences under identical environmental conditions. Most neglected, said Hull, is the relationship between the two approaches. Although Hull acknowledged environmental and historical sources of individual differences, his main concern was with individual differences that are innate and constitutional. His approach, however, was applicable to both sources. As is known, he adopted the point of view of the natural sciences, of physics in particular, where a scientific law is expressed as an equation of a particular form, and the constants in the equation are determined by observed conditions that vary with individual events but do not change the general form of the law. Hull's notion was that individual differences find expression in these constants. Many years later, a few psychologists followed up Hull's notions that individual differences influenced learning equation parameters (Noble, Noble, & Alcock, 1958; Reynolds & Adams, 1954; Spence,

1956, 1960; and Zeaman & Kaufman, 1955). This small amount of work represents a major part of the attention paid by learning theories to individual differences. In contrast, however, at least two approaches to the study of behavior attack the problem of individual differences in learning by attempting to develop techniques that produce lawful individual functions. This is the procedure adopted by Skinner (1938; Ferster & Skinner, 1957), and described in detail by Sidman (1960) in his book on the tactics of scientific research. In a different way, it is also the approach being employed by recent information-processing, computer simulation approaches to the analysis of complex cognitive tasks (Reitman, 1965; Simon & Newell, 1964).

The history of work on learning and individual differences shows clearly the dearth of basic information required for attacking certain critical problems in the design of instruction. The basic problems resolve around issues inherent in adapting educational alternatives (learning conditions) to individual differences at the beginning of a course of instruction and those that appear during learning. Because of the relative insularities of the psychometric field and learning theory, no base of research information and theory is readily available. A major inadequacy of the factor-analytic psychometric approach is the lack of a theoretical framework for the selection of reference tests and learning measures. Global notions of general intelligence are obviously no longer useful scientific concepts for describing learner characteristics because such global measures tend to neglect and obscure specific individual differences. Rather, what is more important

for instruction is to determine initial patterns of ability and competence that interact with learning. In the experimental and theoretical study of learning, resistance to discovering what may be hidden in error variance needs to be overcome. Unique factor variance, if it exists, needs to be examined and accounted for, not only in terms of error, but also in terms of what implications it may have for learning and instruction. As has been indicated [Ed.: cross-reference to latter part of Cronbach's chapter.], learner-treatment interactions must be sought in experiments that study the learning effects of various instructional treatments. Examination of ordinal and disordinal interactions provides the data upon which learning experiences that are adaptive to individual differences can be designed. Increased attention must be paid to initial baseline characteristics of the learner prior to experimental treatment, and statements of principles of learning need to incorporate parameters reflecting individual differences.

Another major contributor to the lack of integration between individual differences and educational alternatives has been the state of educational practice itself. While educators have recognized the need for adapting instruction to individual differences, and various track systems have been devised, the degree of adaptation has never been enough to force answers to the underlying problem of interactions between individual differences and educational alternatives. However, new approaches to individualizing education are being attempted. The problems for instructional design that these new approaches raise will influence both educational practice and the underlying research and knowledge.

Instructional Models

The purpose of measurement for instruction can best be indicated by a particular model for an educational system since different patterns of instruction have different measurement requirements. In general, the model should illustrate that the educational process is concerned with behavioral change and that instruction provides the conditions to foster the processes by which change takes place. Teaching always begins with a particular behavioral state, assesses the characteristics of this state, and implements instructional procedures accordingly; assessment of the changing state of the learner provides information for further use and allocation of instructional methods and resources. Guidance of the instructional process can take place by the student, the teacher, or an automaton. The model should further evidence that an educational system should permit the exercise of individual talents and offer the opportunity for students to develop and excel at every level of ability. It is therefore necessary for an educational system to provide for individualized treatment of students. Educators have been aware of this necessity, and their concern with adapting to the needs of the student is a familiar theme which provides the justification for many current educational innovations (Heathers, 1969).

Several major patterns of adapting to individual differences can be identified in education if one examines past and present educational practices and examines future possibilities (Cronbach, 1967). These patterns can be described in terms of the extent to which educa-

tional goals and instructional methods have been varied for the handling of individual differences as they appear in the school. One pattern occurs where both educational goals and instructional methods are relatively fixed and inflexible. Individual differences are taken into account chiefly by dropping students along the way. The underlying rationale involved is that every child should "go as far as his abilities warrant." However, a weeding-out process is assumed which is reached earlier or later by different individuals. With this pattern, it is also possible to vary "time to learn" required for different students. When this is carried out, an individual is permitted to stay in school until he learns certain essential educational outcomes to a specified criterion of achievement. To some extent, this latter practice is carried out in the old policy of keeping a child in the first grade until he can read his primer and in the more recent nongraded primary unit which some children complete in three years, and some in four.

A second pattern of adaptation to individual differences is one in which the prospective future role of a student is determined, and depending upon this role, he is provided with an appropriate curriculum. When this system is in operation, students are channelled into different courses such as academic courses, vocational courses, or business courses; vocationally oriented students get one kind of mathematics and academically-oriented students get a different kind of mathematics. Adapting to individual differences by this pattern assumes that an educational

system has provision for optional education objectives, but within each option the instructional program is relatively fixed.

In a third pattern of adapting to individual differences, instructional treatments are varied. Different students are taught by different instructional procedures, and the sequence of educational goals is not necessarily common to all students. This pattern can be implemented in different ways. At one extreme, a school can provide a main fixed instructional sequence, and students are branched from this track for remedial work; when the remedial work is successfully completed, the student is put back into the general track. At the other extreme, there is seemingly the more ideal situation. A school carries out an instructional program that begins by providing detailed diagnosis of the student's learning habits and attitudes, achievements, skills, cognitive style, etc. On the basis of this analysis of the student's characteristics, he is guided through a course of instruction specifically tailored to him. Conceivably, in this procedure, students learn in different ways, e.g., some by their own discovery and some by more structured methods.

In light of the current experimentation in schools on procedures for adapting to individual differences, it seems likely that in the near future, patterns falling between these two latter extremes will be developed and adopted by many schools. The quality of the various systems developed will depend upon the answers to many questions of research and practical implementation. Particularly, the difficult question of

the interaction between the characteristics of a student as a particular point in his learning and appropriate methods of instruction is raised for intensive study. Proof will have to be forthcoming that the instructional methods devised for adapting to individual student differences result in significantly greater attainment of educational goals than less intricate classroom practices or classroom practices where the average best method is employed.

The Instructional Model Considered in This Chapter

At the present time, it seems possible to develop educational methods that are more sensitive to individual differences than our procedures have been in the past. Educational systems for accomplishing this will no doubt take many forms and have many nuances as they are developed. The general components of one model are presented here as a basis for examining the measurement and evaluation tasks that it demands. In terms of the three educational patterns of individual difference adaptation described above, it would seem that this model falls somewhere between the extremes of the third pattern, that is, between remedial branching and unique tailoring. It should be pointed out that in an educational pattern adaptive to individual differences, measurement and evaluation tasks arise because certain operations require data and information for decision making. These operations can be categorized into the following six components: (Glaser, 1970)

1. <u>Outcomes of learning are specified in terms of the behavioral manifestations of competence and the conditions under which it is to be exercised.</u> This is the platitudinous assertion of the fundamental necessity for describing the foreseeable outcomes of instruction in terms of certain measurable products and assessable student performance, while at the same time keeping in mind that what is easily measured is not necessarily synonymous with the goals of instruction. In addition, analysis and definition must be made of the performance domain intended to be taught and learned. The "structure" of the domain is specified in terms of its subgoal competencies and possible paths along which students can progress to attain learning objectives.

2. <u>Detailed diagnosis is made of the initial state of a learner entering a particular instructional situation.</u> A description of student performance characteristics relevant to the instruction at hand is necessary to pursue further education. Without the assessment of initial learner characteristics, carrying out an educational procedure is a presumption. It is like prescribing medication for an illness without first describing the symptoms. In the early stages of a particular educational period, instructional procedures will adapt to the findings of the initial assessment, generally reflecting the accumulated performance capabilities resulting from the long-term behavior history of the learner. The history that is specifically measured is relevant to the next immediate educational step that is to be taken.

3. Educational alternatives are provided which are adaptive to the classifications resulting from the initial student educational profiles. These alternative instructional procedures are selectively assigned to the student or made available to him for his selection. They are available through the teacher and/or through materials or automated devices with which the student works.

4. As the student learns, his performance is monitored and continuously assessed at longer or shorter intervals appropriate to what is being taught. In early skill learning, assessment is quite continuous. Later on, as competence grows, problems grow larger; as the student becomes increasingly self-sustaining, assessment occurs less frequently. This monitoring serves several purposes: It provides a basis for knowledge of results and appropriate reinforcement contingencies to the learner and a basis for adaptation to learner demands. This learning history accumulated in the course of instruction is called "short-term history" and, in addition to information from the long-term history, provides information for assignment of the next instructional unit. The short-term history also provides information about the effectiveness of the instructional material itself.

5. Instruction and learning proceed in a cybernetic fashion, tracking the performance and selections of the student. Assessment and performance are interlinked, one determining the nature and requirement for the other. Instruction proceeds as a function of the relationship between measures of student performance, available instructional alternatives, and learning criteria that are chosen to be optimized.

The question of which criteria are to be optimized becomes critical. Is it retention, transfer, the magnitude of difference between pre- and posttest scores, motivation to continue learning including the ability to do so with minimal instructional guidance, or is it all of these? If tracking of the instructional process permits instruction to become precise enough, then a good job can be done to optimize some gains and minimize others unless the presence of the latter gains is desired, expressed, and assessed. The outcomes of learning measured at any point in instruction are referenced to and evaluated in terms of competence criteria and the values to be optimized; provision is always made for the ability of humans to surpass expectations.

6. The system collects information in order to improve itself, and inherent in the system's design is its capability for doing this. A major defect in the implementation of educational innovations has been the lack of the cumulative attainment of knowledge, on the basis of which the next innovation is better than the one that preceded it.

Given that the changing trends in education will lead to an instructional model somewhat like that just described, the following sections of this chapter consider the implications for the nature of measurement and evaluation procedures.

Analysis and Definition of Performance Domains

In an educational system, the specification and measurement of the outcomes of learning in terms of observable human performance determine how the system operates. Vague statements of the desired educational

outcomes leave little concrete information about what the teacher and the student are to look for and what the designers of the system are to strive to attain. Furthermore, performance standards specified in advance need not impose conformities nor stifle freedom of inquiry. Interaction between the specification of outcomes and instructional procedures provides the basis for redefining objectives. The need for constant revision of objectives is as inherent in a well-designed educational system as is the initial need for defining them. There is a sustained process of clarifying goals, working toward them, evaluating progress, reexamining the objectives, modifying instructional procedures, and clarifying the objectives in the light of evaluated experience. This process should indicate the inadequacies and omissions in a curriculum The fear of many educators that detailed specification of objectives limits them to "trivial" behaviors only--those that can be forced into measurable and observable terms--is an incorrect notion. Rather, one should think of them as amendable approximations to our ideals. For example, if complex reasoning and open-endedness are desirable aspects of human behavior, then they need to be recognizable and assessable goals. Failure to state such goals or specification of them in a vague and general way detracts from their being seriously considered as attainable, and may force us to settle for only what can be easily expressed and measured.

The analysis and classification of behavior to be learned is an increasingly prominent feature in the psychology of learning, being fostered both by experimental and theoretical requirements and by attempts at practical applications (Bruner, 1964; Gagné, 1965a,b;

Glaser, 1962; Melton, 1964; Miller, 1965). This trend has come about because all-inclusive theories and schools are no longer major psychological influences and have been replaced by more miniature systems resulting from the analysis of certain behavioral processes and classes of behavior. The working assumption is that the various classes of behaviors that human beings display have different characteristics that need to be specifically analyzed. The implication of this for the analysis of instructionally relevant performance domains is that school learning must be analyzed both for its knowledge content and also its behavioral repertoires.

The increasing movement of individuals between laboratory study and educational problems is contributing to the need for behavior analysis. In the laboratory, a task performed by a subject has special properties built into it for particular scientific interests; the task is so designed that its properties are clear enough for experimental investigation. In contrast, the behavior presented by school learning is not designed for the laboratory and needs to be analyzed so that it can be subjected to study. The necessity for this kind of "task analysis" adds a new requirement to the study of learning and instruction, e.g., recent work in psychology on taxonomies, behavioral categories, and the analysis of behavioral processes (Gagné, 1965a; Melton, 1964; Reitman, 1965; Simon & Paige, 1966). In education, this concern has recently stimulated work on "behavioral objectives" and the definition of educational tasks. Techniques for the analysis of performance and for the derivation of assessment procedures based on these analyses are very

much in the early stages of development, and at the present time this is a growing area of activity among learning and educational psychologists (Gagné, 1970; Gibson, 1965; Glaser, 1962; Hively, 1966a; Kersh, 1965; Schutz, Baker & Gerlach, 1964). Increasingly, there will be more formal analyses of the way in which the content and psychological processes inherent in school learning influence and determine the nature of measurement and instruction.

Subject-Matter Structure and Component Task Analysis

Prominent in the analysis of performance domains is the concern with the structure of the subject matter (e.g., Bruner, 1964; Gagné, 1962; Taba, 1966). As educational tasks or goals are analyzed, they imply a series of subgoals through which instruction must proceed. The arrangement of these subgoals is a function of the subject matter being taught, the approach of the course designer to the subject matter, and also the way in which the student elects, or his performance advises, that instruction should proceed. Different students may follow different paths through the subject matter so that for any particular individual, some subgoals may be omitted, added to, recombined or rearranged. Subgoals provide nodes at which information about performance can be obtained and instructional decisions can be made. There are few techniques available to the analysis of learning tasks and their structure. One procedure that seems most promising is the procedure developed out of Gagné's work on "learning hierarchies" (Gagné, 1962, 1968; Gagné & Paradise, 1961; Gagné and others, 1962). The term "learning hierarchy"

refers to a set of component tasks or performances leading to a particular instructional objective. These component tasks have an ordered relationship to one another. Beginning with a statement of some "terminal" objective of instruction, the attempt is made to analyze this terminal performance into component tasks in a structure such that lower level tasks generate positive transfer to higher level ones. The set of ordered performances forms a hierarchy which can assist in the design of instruction and its assessment.

- - - - - - - - - - - - - - - - - - - -

Insert Figures 1, 2, and 3 about here

- - - - - - - - - - - - - - - - - - - -

Figure 1 reproduces one of these hierarchies pertaining to the addition of integers (Gagné and others, 1962). In the framework of instruction in "modern math," children learn two distinguishable terminal capabilities: One of these, shown on the right, is simply finding sums of positive and negative numbers; a second, shown on the left, constitutes a demonstration of the logical validity of adding any pair of integers, using the properties of the number system to effect this demonstration. For both these tasks, an analysis revealed a set of subordinate capabilities shown in the figure, some in common and some not in common, ranging down to some relatively simple skills which the children were presumed to possess at the beginning of instruction. Figures 2 and 3 show hierarchies of less complex behavior developed with kindergarten children which are somewhat easier to follow (Resnick & Wang, personal communication). In Figure 2 the terminal behavior is counting a movable set of objects; in Figure 3 the terminal behavior is

the capability to place an object in the appropriate cell of a two-dimensional matrix. In each of these two figures the row of double-lined boxes connected by arrows shows the behavioral sequence that accomplishes the terminal performance. The boxes below this show the hierarchical skills leading to this performance sequence. The analysis of learning hierarchies, or component task analysis, begins with any desired instructional objective, behaviorally stated, and asks in effect "to perform this behavior what prerequisite or component behaviors must the learner be able to perform?" For each behavior so identified, the same question is asked, thus generating a hierarchy of objectives based on testable prerequisites. The analysis can begin at any level and always specifies what comes earlier in the curriculum. The importance of the backward analytic procedure for instruction is that it provides a method for identifying critical prior behaviors--behaviors whose absence may be not only difficult to diagnose but also may be significant impediments to future learning. In practical applications, a component task analysis can stop when the behaviors identified are the ones that the course designer believes can be safely assumed in the student population. Thus, this kind of analysis attempts to provide ordered sets of tasks for inclusion in a curriculum and also to specify the skills a student needs to successfully enter a curriculum.

The kinds of performances identified in this manner are not only generated by the logic of the subject matter but also by the psychological structure of the subject matter, psychological structure being roughly defined, in this context, as an ordering of behaviors in a sequence of prerequisite tasks so that competence in an early task in the sequence

facilitates the learning of later tasks in the sequence. The relationship between tasks is hierarchical in the sense that competence at a higher level implies successful performance at lower levels. When analyzed in this way, it may not always be the case that the logical subject-matter relationships in a knowledge structure defined by scholars in the field are the same as the described psychological structure (Glaser, 1962; Suppes, 1966). In the case where one works with task hierarchies for which there is no established subject matter organization, such as the kind of behavior that might be taught to four- or five-year-olds, the nature of the structure of the component tasks is an interesting psychological problem (Resnick, 1967; Resnick & Wang, 1969).

A persistent question that is raised concerns how much of education can be analyzed into hierarchical structures. At this stage of development of instructional design techniques, the answer to the question is very much an open experimental matter. The technique has hardly been explored. Three things should be pointed out, however. First, it should be recognized that hierarchies or structures that might be developed for the more complex behaviors need not be unique. That is, it may well be that several such hierarchies exist, each of which is "valid" with different kinds of learners, but none of which taken singly is valid for all learners. Second, the analysis of learning objectives into component and prerequisite behaviors does not guarantee an immediately complete and viable structure and sequence. As is pointed out below, such hierarchies stand very much as hypotheses subject to empirical investigation. Third, regardless of the precision and specificity with which learning sequences are identified, in actual practice

there is always a functioning sequence. If one is "teaching" a complex behavior, he must begin somewhere and proceed through some sequence of steps. He, thus, has at least an implicit or intuitive structure and sequence within which he operates. The point here is that techniques such as employed by Gagné and by Resnick, for example, provide one means of making explicit the behaviors to be learned and the sequence in which these behaviors might be acquired. It would appear that as these behavioral analysis techniques are improved, much more of the content and process of school subject matter can be analyzed for the purpose of instruction.

Hierarchy Validation

Once analyzed, the hierarchical analysis stands as an hypothesis of ordering that requires data to test its validity. If tests are developed for each of the component tasks described, then data are obtained by which patterns of responding to the subordinate tasks can be ascertained. Indices, somewhat like those obtained in a Guttman-type scale analysis, can be computed to determine the sequential dependencies in the hierarchy (Resnick and Wang, 1969). In contrast to a typical simplex structure, a hierarchical analysis usually presents an intricate tree structure for which new measures of branching and ordering need to be devised. Validation of a hierarchy also can be carried out experimentally by controlled transfer experiments which determine the facilitation in the acquisition of higher ordered tasks as a function of the attainment of lower ones. The empirical tryout of the hypotheses represented by a task hierarchy seems to be an important endeavor for

instructional design. Suggestions about how determinations of hierarchy validity might be made have been discussed in preliminary papers by Gagné (1968), Resnick (1967), and Resnick and Wang (1969). One example is a study by Cox and Graham (1966) using elementary arithmetic tasks. They investigated a task ordering used for instruction, showed how an initially hypothesized ordering might be improved and suggested a revised order that might be more useful to consider in designing the curriculum.

What kinds of information do such structures provide for the design of instruction? The basic implication is that no objective is taught to the learner until he has, in one way or another, met the prerequisites for that objective. However, the prerequisite learnings can be attained in a variety of ways. They can be learned one at a time or they can be learned many at once in large leaps. The instructional process would seem to be facilitated by continuous identification of the furthest skill along the hierarchy that a student can perform at any moment; or if a student is unsuccessful at a particular objective, by determining the most immediate subobjective at which he is successful. The hierarchies as they are derived indicate only the relation of subordination or sequential performance capability. They do not necessarily specify instructional procedures, i.e., how tasks should be learned or what tasks should be taught at the same time. Each analysis says what behaviors are to be observed and tested for, even though it may take a significant amount of instruction to get from one component task to another. As a result, essential information is provided with respect to assessing performance, since the instructor or instructional

device is told what observations are relevant to determining the status of learned performance. A hierarchical analysis provides a good map on which the attainment, in performance terms, of an individual student may be located. The uses of such hierarchies in designing a testing program for a particular instructional system are discussed below.

Placement, Diagnosis, and Assignment to Instructional Treatment

The model of adaptive, individualized instruction outlined previously points to the necessity for specifying foreseeable instructional outcomes and for designing sequences of instructional subgoals that are compatible with the structure of the subject matter and that facilitate attainment of these outcomes. These specified sequences and hierarchies can be considered as a kind of "curricular lattice" through which the progress of individual students can be assessed in their attainment of the instructional goals. If adaptive instruction is at all effective, both the rate and manner of progress through the curriculum sequence will vary from individual to individual. The purpose of this section is to examine the particular measurement requirements involved.

Initial Placement Testing

To facilitate discussion, schematic representations of two types of hierarchical sequences are illustrated in Figure 4. Briefly, the lettered boxes in these illustrations represent instructionally relevant

- - - - - - - - - - - - - - -

Insert Figure 4 about here

- - - - - - - - - - - - - - -

behaviors that are prerequisite to each other. Thus, in the linear sequence, "A" is prerequisite to "B," "B" is prerequisite to "C," etc. In this sequence, "D" represents the terminal instructional outcome for this segment of the instructional sequence. The boxes in the "tree-structure" sequence have a similar relationship, with the exception that parallel columns of boxes are considered to be sequentially independent of each other from a learning sequence point of view. Thus, behaviors "A" and "B" are both considered prerequisite to "D," but "A" and "B" are not prerequisite to each other. Similarly, "D," "E," and "F" are all prerequisite to "G" (the terminal instructional outcome for this sequence), but the temporal sequence of instruction is not specified. Thus, "E" may be learned before "D," "F" learned before "D," etc., but "C" must be learned before "E."

With respect to the individualization of instruction, such a hierarchical specification provides a map on which an individual student may be located before actual instruction begins (i.e., before providing the learning experiences so that the learner may acquire the next sequence of behaviors). Thus, given that little is known about an individual learner who is to acquire the terminal curriculum objective of the sequence, the first decision that must be made about him answers the question, "Where in this sequence of learning experiences should this individual begin his study?" The problem is to locate or place the student with respect to his position in the learning sequence. This first decision, or placement decision, specifies the initial requirements for a testing program designed to facilitate the adaptation of instruction to the individual learner. At this point, the information required of

measuring instruments with respect to a given segment of the instructional sequence is primarily achievement information. These tests provide information concerning the knowledge and skills already possessed by the individual before he begins an instructional sequence. The term "placement test" in this discussion will be reserved for the type of test that provides this kind of information--namely, long-term achievement information that is specifically obtained to facilitate the initial placement decision. It should be noted that the use of the terms "placement" and "placement decision" is somewhat different from the use of those terms in Chapter 15 [Ed.: cross-reference to Cronbach's chapter]. Although here and in Chapter 15 (pp.   ) the concern is with making decisions about all examinees (i.e., there is clearly no screening-out or selection decision), the discrepancy between the two uses of the terms follows from the notion of treatment allocation. That is, at this point in the instructional decision-making process, one is assuming that all of the students being measured by the "placement test" need to be located at some point in the given curriculum sequence and that the decision has not been made concerning the teaching technique (i.e., the instructional treatment) to which an individual is to be assigned in order that he may acquire the next sequential behavior. This latter decision is called a "diagnostic decision" in the discussion below (pp.36-42), and it would seem that some of the statistical characteristics of those tests described in Chapter 15 [Ed.: cross-reference to Cronbach pp. 148 ff.] are more applicable to these latter (diagnostic) instruments. If one either is experimenting with an instructional sequence, or has several viable sequences leading to the same terminal instructional goal to which an

individual may be allocated, then such procedures as outlined in Chapter 15 are important for examining test validities. As an example, consider the two versions of the instructional sequence illustrated at the top of Figure 5. Suppose both were viable sequences for different kinds of students. Suppose one had a predictor test, administered it to a

---

Insert Figure 5 about here

---

group of students, and assigned the students one of the two sequences at random. Then if the regression functions of the outcome measure on the predictor variable appeared as in (i) of Figure 5, one would have some evidence to conclude that Sequence I is a better sequence overall regardless of scores on the predictor test. On the other hand, if the regression functions appeared as in (ii) of Figure 5, one would assign Sequence I to all those who had $Z > Z_i$ on the predictor test, and Sequence I rearranged to all others. However, one would still need to locate a pupil within the particular sequence allocated in order to maximally adapt instruction to individual needs. In this chapter, it is this latter type of decision that will be called a "placement decision."

Achievement information obtained in this way is specific to a particular curriculum sequence, to each prerequisite instructional objective within a given sequence, and to a learner's performance in relation to the given sequence and its prerequisites. Thus, tests designed to provide information for placement decisions in an adaptive instructional system must be constructed with a particular curriculum map in mind. It appears impossible to employ a test based on a vaguely

defined domain of content to provide the information that is required to make an adaptive placement decision of the type considered here. Further, to be useful in placing an individual learner, these tests must yield more than a single, global score reflecting achievement over the entire domain of instruction. Information must be provided concerning the specific knowledge and skills already mastered, partially learned, or not yet mastered by the individual learner. Such placement tests also must provide information about an individual learner's performance which is referenced to the curriculum sequence with which he is faced. This means that the information provided by these placement tests must be accessible to the placement decision-maker in a criterion-referenced form, rather than in a norm-referenced form. For example, in a given group, Johnny's score on a test designed to measure a particular instructional objective may be at the 99th percentile; yet he may well have to be given instruction on the objective. This is so because percentile ranks and, in general, norm-derived scores, are referenced to the group and not referenced to a curriculum sequence as defined here.

It is probable that in situations where little is known about an individual learner's performance and where the curriculum sequence consists of a large number of instructional objectives, a single placement test cannot provide reliable and efficiently obtained information. In certain instructional systems which have attempted adaptive individualization of instruction, an entire curriculum area (such as, elementary mathematics) is structured and sequenced, and placement testing is sequentially performed. For example, Cox and Boston (1967), reporting on the testing procedures employed with Individually Prescribed Instruction

(Glaser, 1967; Lindvall & Bolvin, 1967), demonstrate the use of a sequential testing procedure. In this situation, elementary school mathematics is sequenced in terms of units of instruction. Within each unit is a sequence of instructional objectives that are to be mastered by an individual learner. Initial placement is accomplished in a two-stage testing procedure. A student new to the system is given a test over a broad range of the curriculum sequence, and scores on the test are referenced to specific units within the sequence. The first decision that is made concerns unit placement; at the second stage of testing, placement is made within the unit sequence, to a specific instructional objective. Stage one, broad-range placement to a unit, need occur only once at the beginning of a course of study. When the student completes an instructional unit, he is given a stage-two placement test for the next sequential unit; thus, he is placed within each successive segment of the curriculum sequence. A similar procedure is reported by Rahmlow (1969) with respect to a series of programmed instruction units in mathematics.

Some of the statistical characteristics and decision rules that are applicable to these placement tests are discussed in detail in the section of this chapter dealing with criterion-referenced tests (pp. 56ff). The test characteristics necessary for this type of placement test, if they are to be efficient measuring instruments, depend heavily on the validity of the proposed curriculum sequence. For example, if there were no extant sequence, it would be necessary to test an examinee on every objective (node or "box") in the curriculum. If there is a viable sequence, however, the situation improves considerably. One could then

devise a sequential testing procedure (see pp. 52-55 concerning branch-testing procedures) in which only some nodes are tested, and passing items on those nodes would indicate that earlier nodes in the sequence would be passed by the examinee as well (because of the hierarchical dependencies which exist).

Such a procedure was employed by Ferguson (1969) in designing a computer-assisted placement test for a unit of instruction in the IPI arithmetic curriculum. The hierarchies with which he worked are presented in Figure 6. Figure 6 represents two sequences of instructional objec-

---

Insert Figure 6 and Table 1 about here

---

tives, for a total of 18 instructional objectives in all. (As shown in Table 1, objective number 3 is the same in both sequences) Each one of the 18 instructional objectives defined a relatively homogeneous domain or universe of test items or test tasks. The problem was to locate an individual at a single "box" or objective in each sequence in such a manner that if he were tested on all those objectives below that location he would demonstrate mastery[3] on the items, and if he were tested on all those objectives above that location he would demonstrate lack of mastery on these items. Ferguson found that the most efficient testing procedure was to begin testing with items of "medium difficulty,"[4] for example, items sampled from the universe defined by Objective 8 in Figure 6. If the pupil demonstrated mastery of this objective he was branched to items dealing with an objective that was more difficult, in this case an objective mid-way between the initial objective tested,

Objective 8, and the terminal or most difficult objective. (In Figure 6, Objective 11 satisfies this condition.) If an examinee failed to demonstrate mastery of Objective 8 he was branched to an easier set of items. (In Figure 6 this would be Objective 6). Testing proceeded until a decision was made about each objective, but each objective was not specifically tested since branching to more difficult objectives implied that easier (or lower) objectives have been mastered without formal testing. When the hierarchy is viable, this latter assumption can be substantiated on the basis of empirical results (Ferguson, 1969).

Assignment to Instructional Alternatives

The specification of the structure and sequence of instructional goals and subgoals is a necessary but not a sufficient condition for the adaptation of instruction to the individual. Hierarchical curriculum sequences, as described here, specify neither the rate nor the manner of progress of the individual learner through the sequence, but do indicate what observations to make in assessing learning. Further information is required to determine to which of the available instructional alternatives (i.e., methods or kinds of instruction) different students should be assigned. In terms of instructional content, the placement of learners at various points in the curriculum sequence according to their placement profile provides certain information about the content of instruction or about how instruction should proceed. However, as has been indicated, this procedure is not sufficient with respect to the process or mode of instruction. In terms of decisions to be made, the information required is that which answers the question, "Given that this student has

been located at a particular point in the curriculum sequence, what is the instructional alternative which will best adapt to his individual requirements and thus maximize his attainment of the next instructionally relevant objective?" Such decisions are in a real sense diagnostic decisions[5], in that diagnosis implies both content and nature of the learning "treatment." In this sense, tests designed to provide this kind of information may be called diagnostic tests. It is probably true that a single test of the conventional type now published and used in the schools will not be able to provide all the data relevant to the instructional technique assignment decisions required in an adaptive instructional system.

On the basis of placement and diagnostic information, assignment decisions are made about instructional alternatives. That is, a student is assigned, guided to, or allowed to select a means of instruction. A fundamental question concerns the nature of the instructional alternatives available. What are they? Where do they come from? How are they developed? On what basis do different instructional treatments differ so as to be adaptive to individual requirements? In presently available conventional educational environments, adaptation takes place on the basis of class grouping and perhaps special work with individual students where this is possible. Certain adaptive mechanisms are left up to the student so that some students have to work harder or spend more time on their homework than others. If a school permits a more individualized setting, then other opportunities for providing different instructional alternatives can be made available. Instructional alternatives can be adaptive to the student's present level of achievement and such aspects

as his mastery of prerequisites, his retention of previous learning, the speed at which he learns including the amount of practice he requires, and his ability to learn in structured or less structured situations. Adaptation to treatments differing in these respects, which are shown to be related to measured aspects of entering behavior, might be able to provide a significant beginning for effective adaptation to individual differences. However, in designing instructional alternatives, it is difficult to know how to use other variables which come out of learning theory (such as requirements for reinforcement, distribution of practice, use of mediation and coding mechanisms, and stimulus and modality variables, e.g., verbal, spatial, auditory, and visual presentation), and more needs to be known about their interaction with individual differences. A study by Rosner and others (1969), for example, indicated that there might be relatively high incidence of clinically significant perceptual-motor dysfunction among both special education and regular classroom pupils. Such individual differences should be examined to determine their relationships to educational outcomes (e.g., early reading) and their importance for designing instruction and instructional materials. Another example might be found in the work by Bormuth (1968). Here the reading difficulty [as determined by the cloze readability scale (1969)] of a passage was examined in relationship to the amount of new information a subject acquired from reading the passage. Preliminary results indicated that passages that were "slightly difficult" for the subject resulted in more acquisition of new information than either "too easy," "just right," or "too difficult" passages. If such findings bear up under cross-validation (both over populations of subjects and curriculum areas),

then this might indicate that written instructional materials, say in social studies, need to be adjusted on an individual basis in order to be maximally effective, i.e., adaptive. Several versions of a text, for example, might be needed. Measures of both the text's readability and the pupil's reading level would have to be taken. Textbook assignment would be differential over students, even though they all would cover the same material. Periodic reassignment of texts to coincide with pupil growth in reading ability would be necessary.

If one assumes that measures of entering behavior can be obtained and that instructional treatments are available, then at our present state of knowledge, empirical work must take place to determine those measures most efficient for assigning individuals to classes of instructional alternatives. The task is to determine those measures with the highest discriminating potential for allocating between instructional alternatives. Such measures should have sharply different regression slopes for different instructional alternatives to be most useful [Ed.: cross-reference to Cronbach's chapter pp. 148 ff.]. As a result of initial placement and diagnostic decisions, the group of students involved is reduced to subsets, allocable to the various available instructional treatments. These initial decisions will be corrected by further assignments as learning proceeds so that the allocation procedure becomes a multistage decision process that defines an individualized instructional path.

In this connection, it is to be pointed out that the usual employment of aptitude test batteries has been to predict scholastic success where the instructional system is relatively nonadaptive. The aptitudes generally measured in education are very much the product of the kind of educational environment in which the aptitude tests have been validated. The basic assumption underlying nonadaptive instruction is that all pupils cannot learn a given instructional task to a specified degree of mastery. Adaptive instruction, on the other hand, seeks to design instruction which assures that a given level of mastery is attained by most students. Such models as that proposed by Carroll (1963) and discussed by Bloom (1969) indicate that aptitude takes on a different meaning in adaptive instruction. Other models of adaptive individualized instruction have also been proposed, for example, the IPI project (Lindvall & Bolvin, 1967) and project PLAN (Flanagan, 1967, 1969).

Adaptive instruction demands a different approach to the prediction of success. If the decision to be made is what kind of instruction to provide the learner, then little information is obtained from the usual kind of aptitude measurement. The behaviors that need to be measured are those which are predictive of immediate instructional success within a particular instructional technique. It can be postulated that if the criteria for aptitude test validation had been immediate learning success rather than some long-range performance criteria, the nature of today's generally accepted aptitude batteries would be quite different. This postulation seems likely since factorial studies of the changing composition of abilities over the course of

learning (Fleishman, 1965, 1967) show that different abilities are involved at the beginning and end of the course of learning. While it may be useful to forecast over the long range, an adaptive instructional model also requires measures which are closely related to more immediate learning criteria, that is, success in initial instructional steps. Current types of measured aptitude may be limited in that they are operationally designed to predict over the long period, given reasonably nonadaptive forms of educational treatment. Evidence for this lack of utility of general psychometric measures with respect to instructional decisions comes from the line of studies dealing with correlations between psychometric variables and learning measures (see earlier section on page 9). The identification of the kinds of "aptitude" variables that can be used to place individuals or to recommend to individuals certain kinds of learning experiences is a vast new area in the field of measurement related to instructional decision making.

As has been indicated, aptitude measures are not the only consideration when individuals are allocated to alternate learning experiences to accomplish the same instructional goals. Another aspect of diagnosis includes the analysis of the errors in student responses. One example of a situation in which errors are analyzed and directly related to instructional treatment is found in a series of tests developed by Nesbit (1966). In arithmetic operations involving the addition and subtraction of fractions, children are first given a relatively broad-range test spanning the topic. Those children who err on any of the items are administered a second test. Their errors on the second test are analyzed and the teacher is provided with both a list of the types

of error committed by each child and a description of the specific instructional activities designed to overcome this error. Thus, not only performance omissions (i.e., lack of mastery on the domain of instructional relevant tasks) are identified, but also performance characteristics (i.e., such as error-type identification) and individualized treatment (i.e., learning activities structured around new tasks to be learned and the child's cause of present difficulty) are provided. Testing activities of this sort are to be encouraged if adaptive instruction is to be realized.

Continuous Monitoring and Assessment of Instructional Outcomes

Under the procedures that seem appropriate for the adaptive instruction model, the student, as he proceeds with his course of instruction, has his performance monitored and assessed at established test and decision points. Achievement measures are obtained similar to those used to assess initial placement; in addition, the opportunity is available for assessment to be made of the student's learning characteristics. (Suggestions for the latter have been mentioned above: learning rate, need for practice, ability to retain previous learning, situations in which he seems to learn best, etc.) This achievement and learning style information is updated as the student progresses and provides the primary information for the decision making required to guide student learning. As this continuous measurement is in effect over a period of time, it would incorporate and supercede initial achievement and aptitude information. If appropriately and subtly done, teaching, instruction and testing would fade into one another. Testing information would be used

for the student, teacher, or automaton to make decisions about future instruction, and to a large extent the evaluative, "course grade" function of testing would be deemphasized.

Achievement measurement in this context is necessarily criterion-referenced measurement. The information obtained from a test tells whether a learning criterion has been achieved, and if it has not, further tells in what respect criterion performance has not been attained. Various levels of criterion mastery are set as the student progresses. Generally, some level of mastery is set by the requirements of the subject matter, the student population, etc. Implicit in the instructional model are defined criteria of competence. The basic task for instruction is to provide the methods that will enable most students in a particular course to attain mastery.

Of unique interest in instructional measurement, as instruction proceeds, are the measurements of learning aptitudes and learning styles that can be made. In today's education, assessments of these kinds are, to a large extent, made by observation and judgment of the teacher--when the teacher has the opportunity to observe, is a good observer, and has the appropriate flexibility to implement the results of these judgments. Probably, these observations and judgments can be significantly improved by providing the teacher with observational instruments and by training the teacher in their use. The significant problem in this context is to develop measures of learning characteristics that are useful in practical instruction. As the student learns, it should be possible to devise learning experiences in which measures are

obtained that provide information to the student and the teacher about the student's learning "style." This is an area in which there has been much lip service and which is done intuitively at the present time. The development of appropriate measurement procedures, which might be called learning process psychometrics, seems to be of critical importance (Cronbach, 1967).

As the student learns, then, information is obtained about how he learns and what he learns; instructional assignments, self-made or teacher-made, take place; and assessment is made of a student's performance at particular decision points. There is a three-way relationship between measures of learning, instructional alternatives, and criteria for assessing performance. Since measures of learning and instructional alternatives are evaluated in terms of how well they assist in helping the student attain educational goals, then the criterion measures become quite critical. Depending upon the measures used, some instructional outcomes will be maximized and others minimized; some kinds of student performance may be minimized inadvertently unless they are expressed and explicitly assessed. In this regard, it seems almost inescapable that we develop more fully criterion-referenced measures, measures that reflect a pupil's performance in relation to standards of attainment derived from a behavioral analysis of the curriculum area under consideration. In addition, serious attempts must be made to measure what has been heretofore so difficult: Such aspects as transfer of knowledge to new situations, problem solving, and self-direction--those aspects of learning and knoweldge that are basic to an individual's capability for continuous growth and development.

Two further points are appropriate here. First, information about learning relevant to an adaptive model should come primarily from the interaction effects generally neglected in studies of learning. As Cronbach and Gleser (1965) have pointed out, the learning experimentalist assumes a fixed population and hunts for the treatment with the highest average and least variability. The correlational psychologist has, by and large, assumed a fixed treatment and hunted for aptitude which maximizes the slope of the function relating outcome to measured aptitude. The present instructional model assumes that there are strong interactions between individual measurements and treatment variables; and unless one treatment is clearly the best for everyone, as may rarely be the case, treatments or instructional alternatives should be differentiated in a way to maximize their interaction with performance criteria. If this assumption is correct, then individual performance measures that have high interactions with learning variables and their associated instructional alternatives are of greater importance than measures that do not show these interactions. This forces us to examine the slope of the regression function in learning experiments, so that this interaction can be evaluated. [Ed.: cross-reference to Cronbach's Chapter]. Intensive experimental research is required to determine the extent to which instructional treatments need to be qualified by individual-difference interactions. The search for such interactions has been a major effort in the field of medical diagnosis and treatment and seems to be so in education (Lubin, 1961).

Second, the continuous pattern of assessment and instructional prescription, and assessment and instructional prescription again, can be represented as a multistage decision process where decisions are made sequentially and decisions made early in the process affect decisions made subsequently. The task of instruction is to prescribe the most effective sequences. Problems of this kind in other fields, such as electrical engineering, economics, and operations research, have been tackled by mathematical procedures applied to optimization problems. Essentially, optimization procedures involve a method of making decisions by choosing a quantitative measure of effectiveness and determining the best solution according to this criterion with appropriate constraints. A quantitative model is then developed into which values can be placed to indicate the outcome that is produced when various values are introduced.

An article by Groen and Atkinson (1966) has pointed out the kind of model that may help for this kind of analysis. There is a multistage process that can be considered as a discrete N-stage process; at any given time, the state of the system, i.e., the learner, can be characterized. This state, which is probably multivariate and described by a state vector, is followed by a decision that also may be multivariate; the state is transformed into the new updated state. The process consists of N successive states where at each of the N-1 stages a decision is made. The last stage, the end of a lesson unit, is a terminal stage where no decision is made other than whether the terminal criteria have been attained. The optimization problem in this process is finding a decision procedure for determining which instructional alternatives to present at

each stage, given the instructional alternatives available, the set of possible student responses to the previous lesson unit, and specification of the criteria to be optimized for the terminal stage. This decision procedure defines an instructional strategy and is determined by the functional relationship between (a) the long- and short-range history of the student and (b) student performance at each stage and at the terminal stage. Figure 7 illustrates this type of N-stage

- - - - - - - - - - - - - - - - -

Insert Figures 7 & 8 about here

- - - - - - - - - - - - - - - - -

instructional process as Groen and Atkinson see its application in computer-assisted instruction. A more general flow diagram is presented in Figure 8. This figure illustrates the instructional stages for the Individually Prescribed Instruction Project. To be made useful for the type of analysis described above, the procedure illustrated by Figure 8 would probably need to be broken down into finer stages.

Groen and Atkinson point out that one way to find an optimal strategy is to enumerate every path of the decision tree generated by the multistage process, but that this can be improved upon by the use of adequate learning models which can reduce the number of possible paths that can be considered. In order to reduce these paths still further, dynamic programming procedures (Bellman, 1957; Bellman & Dreyfus, 1962), might be useful for discovering optimal strategies and hence for providing a set of techniques for reducing the portion of the tree that must be searched. This technique involves the maximization or optimization of the utility of a sequence of N decisions (or stages of instruction).

This is accomplished by employing a mathematical function that depends on the maximized utility of the $(N-1)^{\underline{th}}$ decision in the sequence. The utility of a sequence may be defined, for example, in terms of a score on a test that is administered at the completion of the $N^{\underline{th}}$ stage of instruction. Thus, at each of the stages in the sequence of instruction, the learner is presented with the types of instruction that will maximize criterion performance. The kind of instruction presented at each $j^{\underline{th}}$ stage of the sequence is determined as a function of the maximized utility of the instructional decision made at the $(j-1)^{\underline{th}}$ stage. This is an interesting approach for instructional theory and psychometrics to consider, although some initial experimentation has not been overwhelmingly successful and, perhaps, discouraging (Groen & Atkinson, 1966).

In order to carry out such an approach, two fundamental efforts are required: First, quantitative knowledge of how the system variables interact must be obtained, and second, agreed-upon measures of system effectiveness must be established. Upon the completion of these steps requiring, respectively, knowledge and value judgment, optimization procedures can be carried out. It has been shown that relative to the total effort needed to achieve a rational decision, the optimization procedure itself often requires little work when these first two steps are properly done (Wilde & Beightler, 1967). Thus, two ever-present tasks must still be confronted: (a) knowledge and description of the instructional process and (b) the development of valid performance measures.

Management of Test-Provided Information

It is apparent from the preceding discussion in this chapter that the type of information required from a comprehensive testing program in an adaptive system of individualized instruction must be easily generated and readily obtainable by the student and the instructor. This means that a measuring, information-providing system must be designed and embedded as a component of the overall instructional system. Once embedded into the system, instruction and testing become less distinct and mutually supporting.

The information that is generated as a pupil progresses through a curriculum sequence must be processed and analyzed in such a manner that decisions that are to be made with it are facilitated. Thus, testing programs designed to provide the information required to make the four kinds of adaptive decisions--initial placement, individual diagnosis, individual monitoring, and outcome assessment--must also make provisions for reporting results in a usable form to students and instructors. It would seem further that the burden of designing and constructing such tests, of processing response data, and of providing preliminary analysis of test data must be handled by someone other than the classroom teacher. If instructional outcomes and available sequences are specified in advance, there appears to be no reason why tests and other information-generating instruments cannot be predesigned and made available to the student and the instructor as needed. That is, tests can be predesigned and coded to particular segments of the curriculum sequence in much the same manner as texts and other instructional

materials are predesigned. Since the model for individualized instruction considered here (see pp. 17-20) provides for the capacity of the system to update and improve itself as more is learned about its operation, tests and other instructional materials can be updated and reintroduced without disrupting the instructional system. The instructor, then, can be freed of his duties as "materials producer" and can better perform his role as instructional decision-maker and individual adaptor.

The individualization of instruction increases the amount of information required by a multiplicative factor equal to the number of individuals being instructed. Traditionally, group-based information has been the primary source of data used in classroom decision making. When all students are working on the same task, the task (e.g., page number, chapter, etc.) is the only bit of information which is needed to characterize the group. On the other hand, when every individual is allowed to progress at his own rate and to work on different tasks, then one needs distinct information about each student (Cooley, 1970). The kind of information required also varies in the two situations. In the group teaching situation, the information emphasis is on what is taught at any particular point in time. When instruction is adaptive, the information emphasis shifts to what was learned by each pupil.

With the increased amount and kind of information that is required for adaptive instruction, it seems almost inevitable that a computer system be integrated with the measurement and instructional system in order to manage the individualized school. Such a management system

has as its goals to increase the effectiveness of the adaptive instructional model and to maximize teacher productivity in operating in the system. Systems for computer-managed instruction have been described by Bratten (1968), Brudner (1968), and Flanagan (1969). One such computer management system is being developed in connection with an individualized elementary school and is described in detail elsewhere (Cooley & Glaser, 1970). In this system, the instructor can interrogate the computer to obtain a variety of information relevant to making instructional decisions. For example, in curriculum sequences blocked off as units, the instructor is able to obtain a listing of all the performance data available for a particular student who has been working in that unit. This would include test data specific to the unit which was collected prior to instruction (placement data); within-unit performance data and test data (monitoring data); and posttest data over the unit after instruction has been completed (evaluative data). Teacher analysis of these data is used to diagnose and prescribe further work for the student. Another example of information that the instructor can obtain from the computer is a listing of the class members showing where in the curriculum sequence each student is working and how long he has been working at that lesson unit. In this manner, the instructor is able to monitor class progress, and to identify quickly students who are working at a particular point in a sequence for an inordinate length of time; such students may need assignment to new instructional materials, small group instruction, personal tutoring, or other modifications of their instructional environment.

Branch-Testing

In recent years, the advance of instructional technology and the introduction of the computer as an instructional device has spurred serious interest among test constructors in real-time alteration of the manner in which tests are administered and scored--that is, on-the-spot adaptation of the sequences of items, number of items, or manner of presentation of items while testing is in progress. In particular, interest has been generated for a procedure known as branch-testing or tailored-testing. In this testing procedure, the item(s) to which an examinee is to respond next is determined by his responses on the preceding item(s). This procedure permits the possibility that each examinee can be administered a different set of items that are best suited to measuring his characteristics. Thus, tests can be considered "tailored" to the individual. Rules for determining which items to administer next are termed branching rules.

It would seem that tests that are administered in this branching or tailoring mode have great applicability to the four general types of testing problems encountered in the instructional model described above. One application of branching previously mentioned concerns placement testing (Cox and Boston, 1967; Ferguson, 1969; Rahmlow, 1969). Another application was mentioned in connection with diagnostic testing (Nesbit, 1966). In this section, the topic of branch-testing is considered somewhat more broadly to examine the flavor of this type of testing procedure and its possible instructional applications. In another section of the chapter (pp. 33-36), the possibility of using sequential analysis techniques (Wald, 1947) with certain types of test items is discussed.

Most studies dealing with the effectiveness of the branch-testing procedure have been concerned with the measurement of mental ability, that is, the location of an examinee on a continuum of a hypothetical variable or trait (examples of such studies include Bayroff & Seeley, 1967 using the AFQT; Angoff & Huddleston, 1958 using the CEEB; and Cleary, Linn, & Rock, 1968 using the SCAT and STEP). The various strategies of types of branching that have been reported can be subdivided into two broad classes (Cleary, Linn, & Rock, 1968): (1) those procedures which employ two distinct test-sections that route an examinee and measure him, respectively; and (2) those that measure and route examinees simultaneously (i.e., without distinct test-sections to route and measure separately). Within each of these classes, various techniques are employed to construct the routing and/or measuring test, thus giving rise to several branching strategies.

Although these various strategies have been enumerated and described in the literature, little work has been done concerning the instructional implication and possibilities of using branch-testing. In a paper entitled "Some Theory for Tailored Testing," Lord (1970) speaks directly to this point.

> It should be clear that there are important differences between testing for instructional purposes and testing for measurement purposes. The virtue of an instructional test lies ultimately in its effectiveness in changing the examinee. At the end, we would like him to be able to answer every test item correctly. A measurement instrument, on the other hand, should not alter the trait being measured. Moreover, ...measurement is most effective when the examinee knows the answers to only about half of the test items. The discussion here [i.e., the test theory of tailored testing] will be concerned exclusively with measurement problems and not at all with instructional testing (page 2).

Lord's paper shows that from the measurement point of view, gains from tailored testing are little except for low ability and high ability examinees. However, as Green (1968) has indicated in commenting on Lord's paper, branching (particularly under computer control) may have advantages: possible substantial savings in testing time; branching from broad areas of the achievement domain to narrow areas for in-depth analysis; measuring more complex behavior; measuring response latencies; sequencing responses; and sequencing items on the basis of what the measure shows--to name a few. Of particular relevance, when an instructional system is considered, is the point Green makes that considerations of measurement per se are wasteful in the overall decision-making process. Failing to consider the interrelationship between measurement and decision-making neglects the importance of deciding what additional data need to be collected before adequate decisions can be made. The integration of measurement into the decision process has been discussed by Cronbach & Gleser (1965) in the context of selection and placement. It has, however, barely been explored with respect to instruction and with assistance from computers.

Branching strategies for instruction are best based on rules determined by a combination of psychological theory and subject matter organization. For example, in a procedure suggested by Gagné (1969) for assessing the learning of principles, one can distinguish between a principle and concepts that make up the principle. A two-stage testing procedure is employed in which the first item set measures whether or not an individual possesses the concepts. If the individual is successful on these items, he is branched to another set which tests whether or

not he has learned the principle. If one tested only the principle and the student's response was inadequate, it would not be known whether the learner (a) did not learn either the principle or the concepts or (b) learned the concepts but not the principle. Another possibility concerns tasks involving use of two or more principles. The two-stage measurement procedure would be able to discriminate between examinees who (a) knew one principle and not the other, (b) knew the second principle but not the first, (c) knew none of the principles, (d) knew all the principles but were unable to put them together, and (e) knew all the principles and could put them together correctly to solve the task.

A further conception of branch testing can include the notion of measuring the process by which a learner solves a test (e.g., Newell & Forehand, 1968). That is, the examinee is given the task and must interact with and interrogate the computer to determine courses of action or to solicit further information necessary to solve the problem or complete the task. These procedures are not new conceptions in testing (see Glaser, Damrin & Gardner, 1954; McGuire, 1968) but the feasibility of such procedures for measurement seem much greater with computer technology. Moreover, significant advances in measurement in an adaptive instructional system will come about not in the notion of increased precision of measuring the same things we currently measure, but as a result of measurement procedures based upon analyses of subject-matter task structure and the behavioral processes involved in performing these tasks.

Criterion-Referenced Testing

Tests that measure instructional outcomes and that are used for making instructional decisions have special characteristics--characteristics that are different from the mental test model that has been successfully applied in aptitude testing work. That there is a pressing need for the development of achievement or performance measurement theory and technique has been pointed out (Ebel, 1962; Cronbach, 1963; Flanagan, 1951; Glaser, 1963) and although preliminary work has begun, no substantial literature is extant. In this section, some considerations in the development of performance tests are discussed by way of stimulating further the work that is required. Of particular significance are the following: (1) the generation of items from statements of educational objectives; (2) interpretation of a test score in terms of test content and performance criteria, as well as in terms of norms referenced to the scores of other examinees; and (3) interpretation of test scores so that they have meaning beyond the performance sample actually assessed and so that test scores can be generalized to the performance domain which the test subset represents.

At the heart of the issue concerning the two types of tests discussed in this section is the matter of deriving meaning from test scores. The score or number assigned to the individual as a result of a measurement procedure is basically inert and must be related semantically to the behavior of the individual who is measured (Lord & Novick, 1968). There are many semantic interpretations that are

possible in educational measurement, but for the most part, educational test authors have concentrated on interpreting the test score of an individual primarily by relating it to the test scores of other individuals. Such interpretations, which have been called norm-referenced interpretations throughout this chapter, have serious limitations when they are employed with achievement tests that are used in instructional systems seeking to be adaptive to the individual. These limitations were discussed in an earlier section. A complete discussion of why such interpretations have come to be so prevalent in educational measurement is beyond the scope of this chapter, but it can be pointed out that the concentration of psychological test theory on trait variability and on the relative differences between individuals; the reluctance of educators to specify precisely their desired goals in terms of observable behavior; the reliance of measurement specialists on the mental test model; and the desire of test constructors to build tests that are applicable to many different instructional systems for a variety of purposes, have contributed in no small part to the development and use of these norm-referenced interpretations.

The type of semantic interpretation of test scores that is required by the system of adaptive individualized instruction described in this chapter may be termed a criterion-referenced interpretation. <u>A criterion-referenced test is one that is deliberately constructed to yield measurements that are directly interpretable in terms of specified performance standards</u>. Performance standards are generally specified by defining a class or domain of tasks that should be performed

by the individual. Measurements are taken on representative samples of tasks drawn from this domain and such measurements are referenced directly to this domain for each individual measured.

Criterion-referenced tests are not designed only to facilitate individual difference comparisons such as the relative standing of an examinee in a norm group or population, nor are they designed to facilitate interpretations about an examinee's relative standing with respect to a hypothetical variable such as reading ability. Rather, they are specifically constructed to support generalizations about an individual's performance relative to a specified domain of tasks. (In the instructional context, such a domain of tasks may be termed a "domain of instructionally relevant tasks." The insertion of the qualifiers "instructionally relevant" serves to delimit the domain to those tasks, the learning of which is the goal of instruction. The term "tasks" includes both content and process.)

When the term "criterion-referenced test" is used (e.g., by Glaser and Klaus, 1962; Glaser, 1963; Glaser and Cox, 1968; Lindvall and Nitko, 1969), it has a somewhat different meaning from the two more prevalent uses of the terms criterion or criterion tests in educational and psychological measurement literature. One of these usages involves the notion that scores on an achievement measuring instrument (X) correlate with scores derived from a second measurement situation (Y), this second situation being, for example, scores on another achievement test or performance ratings such as grades. With this usage, the Y-scores are often termed criterion scores and the degree to which

the achievement test approximates, or relates to, the criterion is often expressed by the product-moment correlation, r_{XY}. Since the achievement test scores have the potential for correlating with a variety of other measures, relationships to multiple criteria are often reported. A second prevalent interpretation of the term criterion in achievement measurement concerns the imposition of an acceptable score magnitude as an index of attainment. The phrases "working to criterion level" and "mastery is indicated by obtaining a score equivalent to 80 per cent of the items correct" are indicative of this type of interpretation of criterion. Often both of these uses of the term criterion are applied to a single measuring instrument: A test may serve to define the criterion to be measured , and students may be selected according to some cut-off score on it.

Norm-Referenced Tests vs. Criterion-Referenced Tests

As Popham and Husek (1969) indicate, the distinction between a norm-referenced test and a criterion-referenced test is not easily made by the inspection of a particular instrument. The distinction is found by examining (a) the purpose for which the test was constructed, (b) the manner in which it was constructed, (c) the specificity of the information yielded about the domain of instructionally relevant tasks, (d) the generalizability of test performance information to the domain, and (e) the use to be made of the obtained test information.

Since criterion-referenced tests are specifically designed to provide information that is directly interpretable in terms of specified

performance standards, this means that performance standards must be established prior to test construction and that the purpose of testing is to assess an individual's status with respect to these standards. Tests constructed for this purpose yield measurements for an individual that can be interpreted without referencing these measurements to other individuals, i.e., a norm-group. This distinction is a key one in determining whether or not a test is criterion-referenced or norm-referenced. Much the same point was made earlier in this volume in a discussion concerning absolute and differential interpretations [Ed.: cross-reference to Cronbach's chapter pp. 11-12].

One source of confusion between the type of test discussed here and the typical achievement test of traditional usage resides in the notion of defining task domains and sampling from them in order to obtain test items. Arguments are often put forth that any achievement test defines a criterion in the sense that it is representative of desired outcomes and that one can determine the particular skills (tasks) an individual can perform by simply examining his responses to the items on the test. The problem is, of course, that in practice desired outcomes have seldom been specified in performance terms prior to test construction. Further, the items that finally appear on a test have typically been subjected to another rigorous sifting procedure designed to maximize the test constructor's conception of what the final distribution of test scores should be like and how the items of the test should function statistically. Ease of administration and scoring are often other determinants of what the final test task will be. As Lindquist (1968) has noted, many valuable test tasks have been

sacrificed through the machine scoreability requirements of current test practices. These and other other test construction practices often lead to tests composed of tasks that tend to distort interpretations about the capabilities of the examinee with respect to a clearly defined domain of performance standards.

The distinction between norm-referenced and criterion-referenced tests can often be determined by examining the specificity of the information that can be obtained by the test in relation to the domain of relevant tasks. Logical transition from the test to the domain and back again from the domain should be readily accomplished for criterion-referenced tests, so that there is little difficulty in identifying with some degree of confidence the class of tasks that can be performed. This means that the task domain measured by criterion-referenced tests must be defined in terms of observable behavior and that the test is a representative sample of the performance domain from which competence is inferred.

Thus, the attainment of "reading ability" can only be inferred to have occurred. The basis for this inference is observable performance on the specified domain of tasks into which "reading ability" has been analyzed, such as, reading aloud, identifying an object described in a text, rephrasing sentences, carrying out written instruction, reacting emotionally to described events, and so on. Criterion-referenced tests seek to provide information regarding whether such kinds of performance can or cannot be demonstrated by an individual learner and not how much "reading ability" an examinee possesses along

a hypothetical ability dimension. What is implied is some analysis of task structure in which each task description includes criteria of performance. This means that within a particular instructional context a test constructor is seldom free to choose at will the type of task he is to include in his test. This has been already delimited by definition of the domain of relevant tasks that describe the outcomes of learning. It also means that a scoring system must be devised that will preserve information about which tasks an individual can perform. Scores such as percentile ranks, stanines, and grade-equivalents preserve norm-group information but lose the specificity of criterion information (Lindvall and Nitko, 1969).

A criterion-referenced test must also be generalizable to the task domain that the specific test tasks represent. One does not have to go very far in a curriculum sequence before the tasks that the learner is to perform become very large. To take a simple example, in an elementary arithmetic sequence, column addition appears relatively early. An instructionally relevant domain might consist of correct performance on all 3-, 4-, and 5-addend problems with the restriction that each addend be a single-digit integer from 0 through 9. The relevant domain of tasks consists of 111,000 addition problems. The measurement problem for criterion-referenced test constructors is how to build a test of reasonable length so that generalizations can be made about which specific problem types an individual learner can or cannot perform. Norm-referenced test constructors do not have such a problem since judicious selection of items will result in variable

scores which spread out individuals, thus allowing one to say, "Johnny can do more than Suzy." The question of what Johnny can or cannot do is left unanswered. Examination of an individual's item responses provides only a tenuous basis for inference when norm-referenced tests are used (Lindquist and Hieronymus, 1964). Yet, if instruction is to be adaptive to the individual learner, this information must be obtained. Is it specific number combinations which trouble Johnny? Is it problems which involve partial sums of a certain magnitude? Is it failure to apply the associative principle to simplify the calculation? These and many more such questions need to be answered in order to guide the instructional process.

The use to which achievement test information is put is another determinant of whether criterion-referenced or norm-referenced tests are needed. Both kinds of tests are used to make decisions about individuals, but the nature of the decisions determines the information required. In situations where there is a constraint on the number of individuals who can be admitted and in which some degree of selectivity is necessary, then comparisons among individuals are necessary and, hence, norm-referenced information is used. On the other hand, in educational situations where the requirement is to obtain information about the competencies possessed by a single individual before instruction can be provided, then criterion-referenced information is needed. Generally, in existing instructional systems that are relatively non-adaptive, admission decisions are made on a group basis and use norm-referenced data. As the feasibility of adaptive, individualized

instruction increases, knowledge of an individual learner's position in the group becomes less important than knowledge of the competencies that the individual does or does not possess. Hence, it is likely that the requirements of educational measurement will be for criterion-referenced information in addition to norm-referenced information.

Item Construction

The major problem involved in constructing items for criterion-referenced tests is the design of test tasks that are clearly members of the relevant domain. In their ideal form, the tasks to be performed are representative samples of tasks that are the objectives of instruction at a particular stage in the instructional sequence. Two points need to be considered here. The first is the place of ultimate vs. immediate instructional objectives and their relation to instructionally relevant tasks. The second is the generation of test items from descriptions of instructional objectives.

Ultimate and immediate objectives. The distinction between and discussion of ultimate and immediate educational objectives were thoughtfully done by Lindquist (1951) in the previous edition of this volume. Such a distinction and its consequences for educational measurement are especially important to note. Educational practice generally assumes that the knowledge and capabilities with which the student leaves the classroom are related to the educational goals envisioned by the teacher. This assumption implies that the long-range goals that the students are to attain in the future are known and that

the behavior with which they leave a particular course actually contributes to the attainment of these goals. What is closer to reality is that the long-term relationship is not very clear between what the student is taught and the way he is eventually required to behave in society or in his job. In contrast to the ultimate goals of education, the immediate objectives consist of the terminal behavior that a student displays at the end of a specific instructional situation. It should be noted that immediate objectives are not defined as the materials of instruction nor as the particular set of test items that have been used in the instructional situation. For example, at the end of a course in spelling one might reasonably expect a student to be able to spell certain classes of words from dictation. During the course, certain of these words may have been used as examples or as practice exercises. The instructor is interested in the student's performance with respect to the class or domain of words as an immediate objective of instruction and not the particular words used in instruction. Thus, to assess a student's performance with respect to the domain, one may also need to consider the relationship between the items in the domain and the preceding instruction (Bormuth, personal communication).

It is this immediate behavior that is the only tangible evidence on which the teacher can operate and by which both the teacher and the student can determine that adequate instruction is being carried out. However, as Lindquist points out, immediate objectives are ephemeral things: Specific content changes with reorganization of subject matter and methods of teaching; and different instructors in the same subject

want to develop generalized understandings in their students, but each may use quite different subject-matter areas, examples, and materials. Nevertheless, specific end-of-course behaviors are learned by students and tested for by instructors, both operating under the assumption that these behaviors facilitate the attainment of ultimate objectives (although many would not wish to judge the effectiveness of an educational system on the basis of attainment of immediate objectives). The immediate objectives, however, do determine the nature of an instructional institution, the way students and instructors act, and the way in which the success of the teachers, students, and institution is evaluated. In this sense, the present discussion is limited to measurement of those behaviors that are under the control of the educational institution and that the student learns or is expected to learn.

The generation of test tasks. The job of the test constructor is considerably simplified if instructional goals and subgoals are initially specified in terms of relevant tasks that the learner can be expected to perform. Those tasks that are relevant to specific stages in the curriculum sequence, such as one of the "boxes" in Figure 4, form the basis for the tasks to be included in criterion-referenced tests. In recent years, the trend in curriculum design has been to state instructional goals and subgoals in terms of behavioral objectives. Statements of behavioral objectives then must be translated into specific test tasks that, when successfully completed by the individual learner, form the basis for the inference that the behavior has been acquired by the learner. As instructional sequences become complex,

this domain of instructionally relevant tasks becomes quite large but, as Hively (1966b) has indicated, they can often be grouped into classes in such a manner that the general form of a class of tasks can be specified.

Recent developments in the analysis of behavior are helpful in analyzing performance into component tasks. For example, learning hierarchy analysis provides one means of distinguishing between components and more complex behavior. Something like Gagné's (1969) suggestion for a two-stage testing operation is required to measure the presence or absence of the complex behavior and then the presence or absence of the underlying prerequisites or components. The essential point is that adequate measurement must provide unambiguous information about the kinds of behaviors that learners can and cannot perform so that instruction can appropriately proceed. Other examples are Hively's (1966a) analysis and Gibson's (1965) analytical experiments of elementary reading behavior that begin to examine the specific components of reading behavior so that the task domain can be identified for teaching and testing purposes. Another interesting approach has been presented by Gane and Woofenden (1968) using a repetitive mechanical task. Their approach is to express performance in terms of an algorithm or flow chart so that not only are the component tasks specified, but also the sequence of performance is presented. As detailed analyses of school subject matters become increasingly prevalent, the test constructor will be able to judge more easily whether a test task is properly a member of the domain of instructionally relevant tasks or is only possibly related to it.

Specification of the domain of instructionally relevant tasks necessitates more than simply giving examples of the desired tasks. It has been suggested that what is needed is a general "item form" accompanied by a list of task generation rules (Hively, 1966b; Hively, Patterson & Page, 1968). An illustration of such "item forms" is reproduced in Figures 9 and 10. Figure 9 presents examples of "item

Insert Figures 9 and 10 about here

forms" for subtraction tasks in arithmetic skills. A title at the left of the table roughly describes a component task of the subtraction domain. A sample item is given in the next column as it would appear on a test. A general form, together with generation rules, given in the next two columns, defines the set of test items that represent the test task. Specifically, the general form and the rules for generating a set of test items has been called by Hively an "item form." A collection of item forms constitutes a domain or universe from which tests and test items may be drawn. Such a procedure as this delimits and clearly specifies the domain of tasks to be learned and the test constructor can then produce test tasks which clearly represent this domain. Judgments can be made relatively easily concerning the "content validity" of the test. Consider the item form in Figure 10 concerned with a specific ability in algebra performance. In this case an item requiring the solution of the inequality $18 \geq 12 - 2|y + 3|$, is _not_ a member of the domain specified by Figure 10 since there is no application of Postulate 2 to $-2|y + 3|$. A similar approach to defining item

tasks has been presented by Osburn (1968). Osburn's presentation attempts to define a general item type and then to further analyze the general type into more specific item forms so that a hierarchical arrangement of test tasks is generated. His suggestion includes the specification of verbal replacement sets as well as the numerical type depicted by Hively's example. Osburn's example of an item form and a verbal replacement set for one of the variable elements of the item form is reproduced in Figure 11. It would seem that provisions for

- - - - - - - - - - - - - - -

Insert Figure 11 about here

- - - - - - - - - - - - - - -

verbal replacement sets such as these might remove much of the "sterility" that might be encountered by a fixed verbal format, while at the same time maintaining a clear link to a general class of items to be included in a particular test.

Bormuth, in a book entitled On Achievement Test Item Theory (in press) develops the idea that tests that are made using current test construction procedures cannot unequivocally be claimed to represent the properties of instruction nor to be objectively reproduceable. He writes:

> The really critical point is that, in the final analysis, a test item is defined as a property of the test writer and not as a property of the instruction. Hence, a score on an achievement test which is made by the procedures currently in use must be interpreted as the students' responses to the test writer's responses to the instruction. Since we have little knowledge of the factors which determine the test writer's behaviors, we must regard the relationship of the student's score to the instruction as being essentially undefinable. Hence, it seems that what is required is a fundamental change in the conception of a test item, of how it is defined, and of how responses to it are described.

The solution Bormuth offers is to suggest that linguistic analysis can be used to make explicit the methods by which items are derived from statements of instructional objectives. Transformational rules (analogous to linguistic transformations) are used to specify definitions of types of items that could be formed. Like the notion of item forms, a reasonable degree of objectivity and replicability is introduced into item construction procedures.

This brief discussion on item construction has indicated some recent developments for consideration by achievement test constructors concerned with creating test tasks that reliably represent instructional objectives. It is apparent, of course, that these techniques could be applied to tests that are other than criterion-referenced. However, further development and the application of such techniques seem essential to the construction of criterion-referenced tests and for the development of achievement testing theory.

Test Construction

When the domain of instructionally relevant tasks has been analyzed and described, specific test tasks must be selected for inclusion on the final form of the test. Item selection and analysis techniques have, of course, been designed with this in mind. The requirements for norm-referenced or group-based item parameters are well known and are treated extensively in the literature. However, as a study by Cox and Vargas (1966) has indicated, traditional item selection techniques are not uniformly applicable for the design of criterion-referenced tests. The issues of item and test parameters are not clear.

For example, many of the item and test statistics employed with norm-referenced tests are dependent on the observed variance of the total test scores. Criterion-referenced tests, on the other hand, when employed in instructional situations may display little variance in total test scores. For example, instruction in many arithmetic skills, by its very nature, does not seek to "spread-out" the examinees, but seeks to reach criterion levels of general competence. If a test were administered prior to instructional treatment and again after instructional treatment, examinee scores on the posttest would show an increase in mean performance and a decrease in performance variation as each student attained skill mastery. In theory, adaptive instruction seeks to assure that all individuals in the population show certain levels of mastery in the instructional domain. Thus, on those instructional tasks where mastery criteria have been established, if posttest items show great variation in difficulty in the population that has been instructed, and items on the posttest are instructionally relevant tasks, then instruction has been inadequate.

For criterion-referenced tests, the empirical estimation of reliability is not clear. As Popham and Husek (1969) indicate, estimates of internal consistency and test-retest coefficients are often inappropriate because of their dependency on total-test score variability. Perfect performance after instruction for all individuals instructed reduces variance-based estimates to zero. Thus, these estimation techniques may be inappropriate when applied in situations that reflect adaptative instruction. Tests used in these circumstances

could be both internally consistent and stable, yet estimates of these indices that are dependent on score variability may not reflect this.

On the assumption that test tasks are samples from the domain of relevant tasks, the problem of ascertaining an individual's status in a task domain might be conceptualized as an item-sampling problem. That is, tasks are sampled and examined in relation to a single individual. The purpose of the test is to determine the proportion of the tasks in the domain that he can perform. Techniques developed for acceptance sampling and sequential testing (for example, see Lindgren and McElrath, 1966 for an elementary discussion) might be investigated for use in this context. For example, if ϕ represents the "true" proportion of incorrectly performed tasks in the domain for an examinee under consideration, the probability function related to accepting the individual as a "master" of the domain (given ϕ) can be specified and, for a fixed observed cut-off score, probabilities of accepting the individual's test-demonstrated performance as evidence for sufficient mastery of the domain can be computed for each true value of ϕ. One could determine risk in the testing situation for both the examinee and the instructor by specifying in advance the proportion of mastery of the domain required before decisions concerning the continuation or termination of instruction are made. That is, specify criterion error proportions ϕ_1 and ϕ_2 such that if the examinee's error proportion $\phi \leq \phi_1$ he has had sufficient instruction relative to the domain, and if $\phi \geq \phi_2$ more (perhaps different) instruction is indicated. The instructor's "risk" would be allowing a learner to terminate instruction on

this particular domain and get on with new instruction. Examinee "risk" would be forcing the student to continue instruction in the domain when he has already mastered it. The results of some preliminary investigations have been presented by Kriewall and Hirsch (1969) in connection with instruction in elementary mathematics.

In situations where the test length, i.e., the number of test items, can vary from person to person, it may be possible to employ the sequential likelihood-ratio test (Wald, 1947). The procedure allows specification of error rates in advance of testing for given "hypotheses" about the proportion of instructionally relevant tasks (test items) that can be successfully completed by the examinee at a given point in time. A discussion of this technique is found in many elementary statistical texts. In achievement testing applications, this procedure would take on the following character: A student needs to be evaluated on a given, relatively large, domain of tasks. The problem is to determine whether the proportion of correctly performed tasks is sufficient to terminate instruction with respect to this domain and to allow him to advance to instruction on a new domain of tasks. If the proportion of correctly performed tasks is not sufficient for mastery, instruction with respect to the domain is to be continued.

The following proportions are specified in advance of testing.

ϕ_1 = the minimum acceptable proportion of tasks mastered in the domain. This proportion is considered the minimum criterion achievement level for mastery of the domain.

ϕ_2 = an alternative proportion of domain tasks mastered below which the criterion achievement level is not obtained (i.e., the maximum proportion correct that will still result in a non-mastery decision).

In the testing situation, ϕ_1 functions as the null hypothesis to be tested against the alternative ϕ_2. Type I and Type II error rates are then specified for classifying the examinee as having mastery or non-mastery. A Type I error occurs when it is decided that a student needs instruction with respect to the domain, when in fact his true proportion of successfully performed tasks is sufficient for mastery. A Type II error is committed when the student is allowed to terminate instruction, when in fact the true proportion of the tasks he can perform is insufficient for mastery. Acceptance and rejection criteria are then established consistent with the Type I and Type II error rates specified. An examinee continues taking the test until a mastery or non-mastery decision can be made. The acceptance and rejection criteria change after each item is attempted and scored; that is, after each item a decision is made to stop testing and declare mastery, continue testing, or to stop testing and declare non-mastery. This procedure was used successfully by Ferguson (1969) in his work on branch-testing. Items were generated by a computer and presented to the examinee via a teletype terminal. This preliminary study indicated that the sequential sampling technique was feasible. It reduced testing time considerably and yielded reliable mastery decisions with respect to the domains sampled.

These techniques seem interesting but certainly need to be explored further, both theoretically and empirically, before they can be recommended as being useful in the instructional context. They have been discussed briefly here primarily to stimulate further inquiry.

Formative Evaluation

The sixth element of the instructional model considered in this chapter states that the system collects information in order to improve itself and that inherent in the system's design is its capability for doing this. Information feedback for this purpose is an essential aspect of increasing rationality in decision-making relevant to the design of educational programs. Of particular significance in this regard is the recent emphasis on "formative" evaluation (Cronbach, 1963; Lindvall, Cox & Bolvin, 1970; Scriven, 1967). Formative evaluation refers to the data provided during the development and design stages of instructional procedures and materials; these data provide the information used for subsequent redesign of instructional techniques. Information provided to the student or to the teacher only for the conduct of ongoing instruction is not formative in this sense, although the term "formative evaluation" has been used to include both kinds of information (e.g., Bloom, 1969a). Formative evaluation, however, can be included in the intermediate stages of development as well as in later stages of continuous improvement and revision. Throughout, formative evaluation focuses on the specific outcomes of various aspects of instruction so that information is provided about the intended or unintended results of these techniques. In its best sense, formative evaluation precludes the one-shot trial of an innovation on the basis of which a decision is made to accept or reject a new instructional program.

This type of formative evaluation is like the high degree of telemetering instrumentation required for the design of new hardware systems. In the early stages of design, a great deal of instrumentation is devoted to measuring and assessing the characteristics of the various functions that the system carries out and their outcomes. As the system's components become more reliable and information is obtained about their effects, less and less excess measurement for evaluation is necessary. At this point, the information required is only that used for the carrying out of normal operations and for possible eventual improvement. As an example, consider an instructional system, such as IPI, in which one aspect of adaptation to individual differences is the writing of a tailored or individual lesson plan for each student for each skill he is to learn. Such a tailored plan is called a prescription. In the initial and intermediate stages of design and development, it is necessary to collect and analyze teacher prescriptions in order to determine if they are indeed individualized and adaptive to students (Bolvin, 1967). This information is then fed back to system developers (research and development personnel) and to teachers as operators of the system. If it is discovered that prescriptions are not individualized, decisions need to be made concerning whether the system or the operators are the cause. That is, do teachers fail to consider relevant student data and existing alternative instructional treatments, or does the system fail to provide the necessary data and alternative instructional procedures? The relationships between the prescriptive component and other components need to be examined as well. For example, does the testing and measurement component provide the necessary data relevant

to adaptive prescriptions? Such considerations are system evaluations which are formative in nature and serve as a basis for future redesign and development. They also serve to temper examination of only ultimate outcomes such as pupil achievement and pupil progress rates.

The formative evaluation implied by the sixth element of the proposed model requires: (a) a planned and specially designed instructional program, (b) goals that are considered as desirable outcomes of the program, and (c) methods for determining the degree to which the planned program achieves the desired goals. Evaluation studies are generated by concern with the discrepancies among stated, measured, and attained goals; with the discrepancies among the stated means for achieving goals and the actual implemented means; and with an analysis of why implemented means have not resulted in expressed goals. Formative evaluation studies attempt to find out why a program or aspects of a program are or are not effective. The answers require detailed analysis of such factors as the attributes of the program itself (e.g., teaching procedures, instructional materials, testing instruments, classroom management practices), the population of students involved, the situational and community context in which it takes place, and the different effects produced by the program (e.g., cognitive, attitudinal, affective, unintended, and positive or negative side effects). Evaluation can take place along many dimensions and in terms of multiple decision criteria such as learning outcomes, costs, necessity for teacher retraining, community acceptance, etc. The information obtained is feedback to the system and serves to redefine or improve it.

Principles and practices involved in evaluation studies have recently been discussed in detail by many writers: by Suchman (1967) with respect to public service and social action programs in general; by Tyler, Gagné and Scriven with respect to curriculum (1967); by an NSSE yearbook with respect to education in general (Tyler, 1969); by Lindvall, Cox and Bolvin (1970) for individualized educational programs in particular; and others. Campbell and Stanley (1963) describe various aspects of the internal validity of educational experiments. Such considerations are important for formative evaluation procedures carried out to yield information relevant to redesign and development since they relate directly to the interpretation of the effects of the instructional procedure. Bracht and Glass (1968) have discussed the external validity of educational studies, "external" being defined as the extent to which an experiment can be generalized to different subjects, settings, and experimenters. These authors present a detailed examination of the threats to external validity that cause a study to be specific to a limited population or a particular set of environmental factors.

Without going into specific procedures and techniques of evaluation studies, certain general aspects especially appropriate to learning and instruction can be mentioned in this chapter.

Long- and Short-Range Objectives

As has been said previously, a significant problem in the evaluation of instructional systems concerns the relationship between means, immediate instructional objectives, and long-range goals. A program

may be unsuccessful for at least two reasons: Either because it was unsuccessful in developing techniques that produced the desired end-of-course goals or because although it was successful in putting a program into operation and in attaining immediate objectives, these objectives were not related to ultimate expressed goals. Seldom is an instructional enterprise in a position to study the relationship between immediate and ultimate objectives. Programs are usually evaluated in terms of the immediate criteria of school accomplishment or possibly accomplishment in the next higher level of education. Concern for some evaluation of long-range goals has been indicated in Project TALENT (Flanagan, 1964) and the National Assessment Study (Frymier, 1967; Tyler, 1966). For the most part, however, formative evaluation studies concentrate on essentially immediate objectives assuming a relationship between immediate and ultimate goals.

Pre-Innovation Baseline Data

The problem of control groups and comparative studies has been extensively discussed in the literature of educational research (e.g., Campbell and Stanley, 1963). Establishing controls in the light of the many interacting factors that influence school settings and populations is a major difficulty in the conduct of evaluation studies. In recent years, particularly in special education, techniques suggested by the work of Skinner have been used with individual children in which the learner is used as his own control. These techniques have been described by Wolf and Risley (1969) and in the context of basic scientific research in behavior by Sidman (1960). It is of interest to consider

these techniques in the context of formative evaluation. An essential aspect of the design used in these studies is the establishment of baselines. The use of baseline logic proceeds by asking the question "Does the instructional treatment substantially affect the baseline rate of the learner's behavior?" The question implies that a change occurs and that sufficient information is obtained to attribute the change to the instructional procedure. For this purpose, measures of relevant aspects of the learner's behavior are obtained prior to the introduction of new instructional techniques. The new techniques are then introduced and change is observed in relation to the previously obtained baseline measures. Assuming that measurement of baseline aspects had been in effect long enough to indicate that the measures were reasonably stable and that the changes after the instructional treatment were significant, it still might be difficult to attribute the change to the specifics of the new instructional procedures. To pin down cause and effect, some form of control comparisons is desirable, and possible designs, in educational settings, that provide sufficient information for making an estimate of change have been suggested by Wolf and Risley (1969). Related also is the discussion by Campbell & Stanley (1963) of the time series experiment and the equivalent time samples design.

The import of employing such techniques as these is that evaluation studies generally have not reported pre-innovation baseline data, and the detailed assessment of the students, teachers, and school environment prior to the introduction of new instructional techniques seems fundamental to effective evaluation.

The Independent Variable

The formative evaluation implied by the sixth element of the model assesses the effect of practices derived from elements one through five. The practices are introduced for the attainment of expressed objectives. Not only must the degree to which objectives are attained be ascertained, but also the effectiveness with which the practices are carried out must be determined. Appropriate values of the dependent variable, i.e., attainment of objectives, it is assumed, will result from effective implementation of the independent variable, i.e., the practices developed to implement the first five elements of the model. However, in most educational studies, more attention is paid to assessing outcomes rather than the adequacies of implementation. Certainly, the latter is a prior requirement. In order to accomplish this, it is necessary for the designers of an instructional program to provide specific criteria that indicate just how the program should function and how specific features of the program should look when the program is in actual operation. A listing of the criteria for the satisfactory functioning of these items provides a checklist for evaluating the degree to which adequate implementation has taken place.

Determining the effectiveness of the independent variable is one major requirement of the instructional model described in this chapter. Assessments of the operation of the program are needed in order to provide information for redesigning and improving its implementation. The other major aspect is whether or not adequate implementation can indeed accomplish program objectives. In reality, in the day-to-day development of instructional programs, the distinction

between these two aspects is not clear. As one assesses whether teaching, materials, equipment, and general school practices are operating appropriately, information is also obtained about how they affect instructional objectives. One usually does not wait to get near-perfect implementation and then proceed to measure instructional outcomes. In the stages of formative evaluation, both aspects proceed together. It is only after some degree of stability is attained and a program has been developed that it seems reasonable to move into a second phase of development. In this second stage, every effort is made to ensure that the implementation criteria are met for the most part, and when they are, goals of the program can be evaluated more definitely. An example of the specification of items in the operation of an instructional program has been described by Lindvall, Cox, and Bolvin (1970) for the program on Individually Prescribed Instruction. Such a specification is geared to evaluating the program's implementation. Basic program operations have been broken down into the following classes: characteristics of instructional objectives, testing procedures, the prescribing of instruction, instructional materials and devices, teacher activities, student activities, and classroom management procedures. Figure 12

- - - - - - - - - - - - - - -

Insert Figure 12 about here

- - - - - - - - - - - - - - -

shows each of these classes of operations in outline form. The operations listed are those that need to be observed and assessed, and for which criteria must be stated, at a particular stage of development of the program, so as to indicate adequate or inadequate implementation.

Such a list of specifications provides the basis for the development of telemetering procedures that are used by instructional developers to monitor the implementation of the independent variables and to determine the internal validity of the results of the instructional techniques.

Particular comment should be made on instructional materials and devices that appear to be a new element for evaluation in present-day instructional programs. Some general principles involved have been described by Lumsdaine (1965) and by Mechner (1965). An examination of the product development process and the training of personnel in the field (Popham, 1967), and examples of its effectiveness have been documented (Flanagan, 1966; Mechner, 1967). The evaluation of materials and devices has many facets that need to be examined, such as: the sequencing and content of instruction, format and packaging, the ability of the student to follow directions for use, the student's ability to manipulate and work with materials and devices of a particular design, and the way in which the teacher employs these techniques. Procedures are being developed for product design and evaluation along a number of lines. For example, with respect to computer-assisted instruction, Bunderson (1970) has described components of a prescriptive model for designing CAI programs. An interesting technique for evaluating material in programmed instructional texts has been described by Holland (1967); and the evaluation hierarchies in specific subject matters have been described by Gagné (1970) and by Resnick and Wang (1969).

In much the same manner as test designers obtain data on test characteristics in order to improve test functioning, data on instructional techniques need to be obtained. Just as the design-trial-redesign cycle has been used in the development of programmed instructional materials, formative evaluation proceeds for educational systems in general. It seems likely that techniques employed for instruction will eventually, where applicable, be developed with the same degree of analysis and documentation as is now done for well received test batteries. The history of evaluation in the testing movement is clear: As tests came to be increasingly used and abused, professional societies stepped in to issue statements of standards for quality control, and schools of education provided courses in tests and measurements for users. At the present time, test producers provide manuals documenting the development and specific utility of the tests under particular conditions and with particular populations. Vis-à-vis the present technology of test construction, design and evaluation with respect to instruction will have to develop its own theories and practices growing out of a convergence of the fields of individual differences, learning, and performance analysis. Some departure will be required in the standard rules of test development and use (Cronbach, 1963).

Sustaining Mechanisms

At the later stages of formative evaluation or following an encouraging evaluation study, a significant concern often is whether or not the effects of the experimental instructional technique will hold up as a continuing state of affairs. One aspect of this is the so-

called "Hawthorne effect." In the classic Hawthorne study (Roethlisberger & Dickson, 1939), an evaluation of a program designed to increase worker productivity found that the specific operational independent variable such as changes in illumination, rest periods, and hours of work were spuriously effective; that is, productivity tended to increase no matter what change was made. The investigators concluded that the actual independent variable causing change was interest and concern on the part of the management. A well executed evaluation study should be able to detect such effects. Factors that result in only the temporary maintenance of effects may be extremely subtle and may not be immediately apparent. The maintenance of effects requires environmental support for the new program. Frequently, when teachers are trained in new curricula and techniques which they bring to their classrooms, conditions are provided in which the new program can proceed, but eventually conventional forces of the environment resume their potency and the innovation is stifled. An example of this is the series of events that followed the introduction of programmed texts into conventional school settings. A study by Carlson (1965) described some of the effects of the lack of a supporting environment for this new instructional technique. One of the unanticipated consequences he described was a restriction of individual differences in learning rate. Although an important anticipated consequence of programmed instruction was that students could be able to learn at their own rates, there were forces operating which minimized the differences in individual rates of achievement. As the program progressed, and as individual students began to vary widely in levels of achievement and rates of progress,

the teacher "corrected" for this by either consciously or unconsciously pacing the students. The output of the fast students was restricted so that the same troublesome point could be explained to a number of students at one time, and the slow students were allowed to have access to programs outside of class time while average and fast students were not allowed extra-class access. This had the net effect of minimizing the range of student progress. In addition, "enrichment materials" were supplied to the fast students which also contributed to a condition of minimum spread. In this and other respects, when programmed instruction materials were introduced into a school for further evaluation, sustaining mechanisms were not provided that would permit the impact of this new instructional technique to result in its anticipated consequences.

Adaptation to Individual Differences

The key issue in instructional systems that attempt to individualize instruction is evaluation of the effectiveness of techniques designed for adapting instruction to individual differences. The instructional model employed as an organizing basis for this chapter attempts to present a set of general requirements for individualizing instruction. However, the success of any model for individualization is limited by certain constraints. If the operational plan is carried out satisfactorily, then the limitations become ones of technical capability and the extent of knowledge about human behavior. This revolves about several basic issues: the extent to which, in any particular subject matter, learning hierarchies or other orderly structures can

be identified and validated; the extent to which individual differences in background and learning characteristics that interact with instructional variables can be identified and measured; and the extent to which alternative instructional techniques and educational experiences can be developed that are adaptive to these measured individual characteristics. These issues are significant areas for basic research in the areas of human performance analysis, the measurement of individual differences, and the functional relationship between these differences and the details of the learning process. The tasks of formative evaluation are to assess technological developments based upon what fundamental knowledge is available, to force improved application, and to provide questions for basic research. The extent to which systems of individualized education are successful in adapting to the nuances of individual differences is a function of this knowledge. The criterion against which systems for individualized instruction need to be evaluated is the extent to which they optimize the use of different measures of behavior and different alternatives for learning in order to provide different instructional paths. It is possible to overdifferentiate and underdifferentiate in adapting to individual differences, and evaluation might indicate that only a relatively few number of paths are more effective in attaining educational goals than a conventional system which teaches to the average student. As more knowledge is obtained, the number of paths available for different individuals will be determined by our knowledge of the relationships between learning, the analysis of learned performance, and measures of individual differences.

References

Angoff, W. H., & Huddleston, E. M. The multi-level experiment: A study of a two-level test system for the College Board Scholastic Aptitude Test. Statistical Report 58-21. Princeton, New Jersey: Educational Testing Service, 1958.

Bayroff, A. G., & Seeley, L. C. An exploratory study of branching tests. Technical Research Note 188, U. S. Army BSRL, June 1967.

Bellman, R. Dynamic programming. Princeton, New Jersey: Princeton University Press, 1957.

Bellman, R., & Dreyfus, S. E. Applied dynamic programming. Princeton, New Jersey: Princeton University Press, 1962.

Bloom, B. S. Some theoretical issues relating to educational evaluation. In R. Tyler (Ed.), Educational evaluation: New roles, new means. Sixty-eighth Yearbook of the National Society for the Study of Education, Part II, 1969, 26-50. (a)

Bloom, B. S. Learning for mastery. In T. Hastings & G. F. Madaus (Eds.), Formative and summative evaluation of student learning. New York: McGraw-Hill, 1969. (in press) (b)

Bolvin, J. O. Evaluating teacher functions. Paper presented at The American Educational Research Association, New York, February 1967.

Bormuth, J. R. Empirical determinants of the instructional reading level. Paper presented at the International Reading Association Conference, Boston, April 1968.

Bormuth, J. R. On achievement test item theory. Chicago: University of Chicago Press, 1969. (in press)

Bracht, G. H., & Glass, G. V. The external validity of experiments. American Educational Research Journal, 1968, 5, 437-474.

Bratten, J. E. Educational applications of information management systems. Santa Monica, Calif: System Development Corporation, June 1968.

Brudner, H. J. Computer-managed instruction. Science, 29 November 1968, 162, 970-976.

Bruner, J. S. Some theorems on instruction illustrated with reference to mathematics. In E. Hilgard (Ed.), Theories of learning and instruction. 63rd Yearbook, Part I. Chicago: National Society for the Study of Education, University of Chicago, 1964. Pp. 306-335.

Bunderson, C. V. The computer and instructional design. In W. Holtzman (Ed.), Computer-assisted instruction, testing, and guidance. New York: Harper & Row, 1970 (in press)

Campbell, D. T., & Stanley, J. C. Experimental and quasi-experimental designs for research. In N. L. Gage (Ed.), Handbook of research on teaching. Chicago: Rand-McNally, 1963. Pp. 171-246.

Carlson, R. C. Adoption of educational innovations. Eugene, Oregon: The Center for the Advanced Study of Educational Administration, University of Oregon, 1965.

Carroll, J. B. A model of school learning. Teachers College Record, 1963, 64, 723-733.

Cleary, T. A., Linn, R. L., & Rock, D. A. An exploratory study of programmed tests. Educational and Psychological Measurement, 1968, 28, 347-349.

Cooley, W. W. Computer-assisted instructional management. Encyclopedia of Education, New York: Macmillan, 1970. (in press)

Cooley, W. W., & Glaser, R. The computer and individualized instruction. Science, 13 October 1969, 166, 574-582.

Cox, R. C., & Boston, M. E. Diagnosis of pupil achievement in the Individually Prescribed Instruction Project. Working Paper 15. Pittsburgh, Pa.: Learning Research and Development Center, University of Pittsburgh, 1967.

Cox, R. C., & Graham, G. T. The development of a sequentially scaled achievement test. Journal of Educational Measurement, 1966, 3, 147-150.

Cox, R. C., & Vargas, J. S. A comparison of item selection techniques for norm-referenced and criterion-referenced tests. Paper read at the National Council on Measurement in Education, Chicago, February 1966.

Cronbach, L. J. The two disciplines of scientific psychology. American Psychologist, 1957, 12, 671-684.

Cronbach, L. J. Course improvement through evaluation. Teachers College Record, 1963, 64, 672-683.

Cronbach, L. J. How can instruction be adapted to individual differences? In R. Gagné (Ed.), Learning and individual differences. Columbus, Ohio: Charles E. Merrill Books, Inc., 1967. Pp. 23-29.

Cronbach, L. J., & Gleser, G. C. Psychological tests and personnel decisions. Urbana, Ill.: University of Illinois Press, 1965.

Duncanson, J. P. Intelligence and the ability to learn. Princeton, New Jersey: Educational Testing Service, 1964.

Ebel, R. L. Content-standard test scores. Educational and Psychological Measurement, 1962, 22, 15-25.

Ferster, C. B., & Skinner, B. F. Schedules of reinforcement. New York: Appleton-Century-Crofts, 1957.

Ferguson, R. L. The development, implementation, and evaluation of a computer-assisted branched test for a program of individually prescribed instruction. Unpublished doctoral dissertation, University of Pittsburgh, 1969.

Flanagan, J. C. Units, scores, and norms. In E. F. Lindquist (Ed.), Educational Measurement, 1962, 22, 15-25.

Flanagan, J. C. The identification, development, and utilization of human talents: The American high school student. USOE Cooperative Research Project No. 635. University of Pittsburgh, Pittsburgh, Pa., 1964.

Flanagan, J. C. The assessment of the effectiveness of educational programs. Mimeograph, May 1966.

Flanagan, J. C. Functional education for the seventies. Phi Delta Kappan, 1967, 49, 27-32.

Flanagan, J. C. Program for learning in accordance with needs. Psychology in the Schools, 1969, 6, 133-136.

Fleishman, E. A. The description and prediction of perceptual-motor skill learning. In R. Glaser (Ed.), Training Research and Education. New York: Wiley, 1965. Pp. 137-175.

Fleishman, E. A. Individual differences and motor learning. In R. M. Gagné (Ed.), Learning and individual differences. Columbus, Ohio: Charles E. Merrill Books, Inc., 1967. Pp. 165-191.

Frymier, J. R. National assessment. In F. T. Wilhelm (Ed.), Evaluation as feedback and guide. Washington, D. C.: Association for Supervision and Curriculum Development and National Education Association, 1967. Pp. 249-259.

Gagné, R. M. The acquisition of knowledge. Psychological Review, 1962, 69, 355-365.

Gagné, R. M. The conditions of learning. New York: Holt, Rinehart & Winston, 1965 (a)

Gagné, R. M. The analysis of instructional objectives for the design of instruction. In R. Glaser (Ed.), Teaching machines and programed learning II: Data and directions. Washington, D. C.: National Education Association, 1965. Pp. 21-65. (b)

Gagné, R. M. Learning and individual differences. Columbus, Ohio: Charles E. Merrill Books, Inc., 1967. (a)

Gagné, R. M. Curriculum research and the promotion of learning. In Perspectives of curriculum evaluation. AERA Monograph Series on Curriculum Evaluation, No. 1. Chicago: Rand McNally and Company, 1967. Pp. 19-38. (b)

Gagné, R. M. Learning hierarchies. Educational Psychologist, November 1968, 6, 1-9.

Gagné, R. M. Instructional variables and learning outcomes. In M. C. Wittrock and D. Wiley (Eds.), Evaluation of instruction. New York: Holt, Rinehart & Winston, 1969. (in press)

Gagné, R. M., & Paradise, N. E. Abilities and learning sets in knowledge acquisition. Psychological Monographs, 1961, 75 (Whole No. 578)

Gagné, R. M., Mayor, J. R., Garstens, H. L., and Paradise, N. E. Factors in acquiring knowledge of a mathematical task. Psychological Monographs, 1962, 76, (Whole No. 526).

Gane, C. P., and Woolfenden, P. J. Algorithms and the analysis of skill behaviors. Industrial Training International, July 1968.

Gibson, E. J. Learning to read. Science, 1965, 148, 1066-1072.

Glaser, R. Some research problems in automated instruction: Instructional objectives and subject-matter structure. In J. E. Coulson (Ed.). Programmed learning and computer-based instruction. New York: John Wiley & Sons, 1962. Pp. 67-85.

Glaser, R. Instructional technology and the measurement of learning outcomes. American Psychologist, 1963, 18, 519-521.

Glaser, R. Adapting the elementary school curriculum to individual performance. Proceedings of the 1967 Invitational Conference on Testing Problems. Princeton, New Jersey: Educational Testing Service, 1967. Pp. 3-36.

Glaser, R. Evaluation of instruction and changing educational models. In M. C. Wittrock & D. Wiley (Eds.), Evaluation of instruction. New York: Holt, Rinehart, & Winston, 1969. (in press)

Glaser, R., & Cox, R. C. Criterion-referenced testing for the measurement of educational outcomes. In R. Weisgerber (Ed.), Instructional process and media innovation. Chicago: Rand-McNally, 1968. Pp. 545-550.

Glaser, R., & Klaus, D. J. Proficiency measurement: Assessing human performance. In R. Gagné (Ed.), Psycholgocial principles in system development. New York: Holt, Rinehart, & Winston, 1962. Pp. 419-474.

Glaser, R., Damrin, D. E., & Gardner, F. M. The tab item: A technique for the measurement of proficiency in diagnostic problem solving tasks. Educational and Psychological Measurement, 1954, 14, 283-293.

Green, B. F. Comments on tailored testing. In W. Holtzman (Ed.), Computer-assisted instruction, testing, and guidance. New York: Harper & Row, 1969. (in press)

Groen, G. J., & Atkinson, R. C. Models for optimizing the learning process. Psychological Bulletin, 1966, 66, 309-320.

Heathers, G. Grouping. Encyclopedia of Educational Research. Fourth Edition. New York: Macmillan, 1969. Pp. 559-570.

Hively, W. A framework for the analysis of elementary reading behavior. American Educational Research Journal, 1966, 3, 89-103. (a)

Hively, W. Preparation of a programmed course in algebra for secondary school teachers: A report to the National Science Foundation. Minnesota National Laboratory, Minnesota State Department of Education, 1966. (b)

Hively, W., Patterson, H. L., & Page, S. A "universe-defined" system of arithmetic achievement tests. Journal of Educational Measurement, 1968, 5, 275-290.

Holland, J. G. A quantitative measure for programmed instruction. American Educational Research Journal, 1967, 4, 87-101.

Hull, C. L. The place of innate individual and species differences in a natural-science theory of behavior. Psychological Review, 1945, 52, 55-60.

Kersh, B. Y. Programming classroom instruction. In R. Glaser (Ed.), Teaching machines and programed learning II: Data and directions. Washington, D. C.: National Education Association, 1965. Pp. 321-368.

Kriewall, T. E., & Hirsch, E. The development and interpretation of criterion-referenced tests. Paper presented at the American Educational Research Association, Los Angeles, California, February 1969.

Lindgren, B. W., & McElrath, G. W. Introduction to probability and statistics. Second Edition. New York: Macmillan, 1966.

Lindquist, E. F. Preliminary considerations in objective test construction. In E. F. Lindquist (Ed.), Educational measurement. Washington, D. C.: American Council on Education, 1951. Pp. 119-158.

Lindquist, E. F. The impact of machines on educational measurement. Bloomington, Indiana: Phi Delta Kappa International, 1968.

Lindquist, E. F., & Hieronymus, A. N. Teachers Manual: Iowa tests of basic skills. Boston: Houghton-Mifflin, 1964.

Lindvall, C. M., & Bolvin, J. O. Programed instruction in the schools: An application of programing principles in "Individually Prescribed Instruction." In P. Lange (Ed.), Programed instruction. 66th Yearbook, Part II. Chicago: National Society for the Study of Education, 1967. Pp. 217-254.

Lindvall, C. M., & Nitko, A. J. Criterion-referenced tests: A rationale for their development and an empirical investigation of their use. Paper presented at the National Council on Measurement in Education, Los Angeles, February 1969.

Lindvall, C. M., Cox, R. C., & Bolvin, J. O. Evaluation as a tool in curriculum development: The IPI evaluation program. AERA Monograph Series on Curriculum Evaluation. Chicago: Rand-McNally, 1970. (in press)

Lord, F. M., & Novick, M. R. Statistical theories of mental test score. Reading, Mass.: Addison-Wesley, 1968.

Lord, F. M. Some test theory for tailored testing. In W. Holtzman (Ed.), Computer-assisted instruction, testing, and guidance. Harper & Row, 1970. (in press)

Lubin, A. The interpretation of significant interaction. Educational and Psychological Measurement, 1961, 21, 807-817.

Lumsdaine, A. A. Assessing the effectiveness of instructional programs. In R. Glaser (Ed.), Teaching machines and programed learning II: Data and directions. Washington, D. C.: National Education Association, 1965. Pp. 267-320.

McGuire, C. H. An evaluation model for professional education--medical education. In Proceedings of the 1967 invitational conference on testing problems. Princeton, New Jersey: Educational Testing Service, 1968. Pp. 37-52.

Mechner, F. Science education and behavioral technology. In R. Glaser (Ed.), Teaching machines and programed learning II: Data and directions. Washington, D. C.: National Education Association, 1965. Pp. 441-507.

Mechner, F. Behavioral analysis and instructional sequencing. In P. Lange (Ed.), Programed instruction. 66th Yearbook, Part II. Chicago: National Society for the Study of Education, 1967. Pp. 81-103.

Melton, A. W. Categories of human learning. New York: Academic Press, 1964.

Miller, R. B. Analysis and specification of behavior for training. In R. Glaser (Ed.), Training research and education. New York: Wiley, 1965. Pp. 31-62.

Nesbit, M. Y. The CHILD program: Computer help in learning diagnosis of arithmetic scores. Curriculum Bulletin 7-E-B. Miami, Florida: Dade County Board of Public Instruction, 1966.

Newell, A., & Forehand, G. On process measurement, computer simulation and theoretical models. In H. H. Harmon, C. E. Helm, and D. E. Loye (Eds.), Computer-assisted testing: Proceedings of a conference on CAT. Princeton, New Jersey: Educational Testing Service, 1968.

Noble, C. E., Noble, J. L., & Alcock, W. T. Prediction of individual differences in human trial-and-error learning. Perceptual and Motor Skills, 1958, 8, 151-172.

Osburn, H. G. Item sampling for achievement testing. Educational and Psychological Measurement, 1968, 28, 95-104.

Popham, W. J. Instructional product development: Two approaches to training. AV Communication Review, 1967, 15, 402-411.

Popham, W. J., & Husek, T. R. Implications of criterion-referenced measurement. Journal of Educational Measurement, 6, 1969, 1-9.

Rahmlow, H. F. A measurement design to facilitate optimum efficiency in diagnosing student needs. Paper presented at the National Council on Measurement in Education, Los Angeles, February 1969.

Reitman, W. Cognition and thought. New York: John Wiley & Sons, 1965.

Resnick, L. B. Design of an early learning curriculum. Working Paper 16. Pittsburgh, Pa.: Learning Research and Developing Center, University of Pittsburgh, 1967.

Resnick, L. B., & Wang, M. C. Approaches to the validation of learning hierarchies. Paper presented at the Eighteenth Annual Western Regional Conference on Testing Problems, San Francisco, May 1969.

Reynolds, B., & Adams, J. A. Psychomotor performance as a function of initial level of ability. American Journal of Psychology, 1954, 67, 268-277.

Roethlisberger, F. J., & Dickson, W. J. Management and the worker. Cambridge, Mass.: Harvard University Press, 1939.

Rosenthal, R., & Jacobson, L. Pygmalion in the classroom. The Urban Review, September, 1968, 3, 16-20.

Rosner, J., Richman, V., & Scott, R. H. The identification of children with perceptual-motor dysfunction. Working Paper 47. Pittsburgh, Pa.: Learning Research and Development Center, University of Pittsburgh, 1969.

Schutz, R. E., Baker, R. L., & Gerlach, V. S. Measurement procedures in programmed instruction. Tempe, Arizona: Classroom Learning Laboratory, Arizona State University, 1964.

Scriven, M. The methodology of evaluation. In Perspectives of curriculum evaluation. AERA Monograph Series on Curriculum Evaluation, No. 1. Chicago: Rand-McNally, 1967. Pp. 39-83.

Sidman, M. Tactics of scientific research. New York: Basic Books, Inc., 1960.

Simon, H. A., & Newell, A. Information processing in computer and man. American Scientist, 1964, 52, 281-300.

Simon, H. A., & Paige, J. M. Cognitive processes in solving algebra word problems. In B. Kleinmuntz (Ed.), Problem solving. New York: Wiley, 1966. Pp. 51-119.

Skinner, B. F. The behavior of organisms: An experimental analysis. New York: Appleton-Century-Crofts, 1938.

Spence, K. W. Behavior theory and conditioning. New Haven, Conn.: Yale University Press, 1956.

Spence, K. W. Behavior theory and learning. Englewood Cliffs, N. J.: Prentice-Hall, 1960.

Stake, R. E. Learning parameters, aptitudes, and achievement. Psychometric Monographs, 1961, No. 9.

Suchman, E. A. Evaluation research: Principles and practice in public service & social action programs. New York: Russell Sage Foundation, 1967.

Suppes, P. Mathematical concept formation in children. American Psychologist, 1966, 21, 139-150.

Taba, H. Teaching strategies and cognitive functioning in elementary school children. Cooperative Research Project No. 2404. San Francisco: San Francisco State College, 1966.

Thorndike, E. L. Educational psychology. Volume III. New York: Teachers College, Columbia University, 1914.

Tyler, R. W. The objectives and plans for a national assessment of educational progress. Journal of Educational Measurement, 1966, 3, 1-4.

Tyler, R. W. Changing concepts of educational evaluation. In Perspectives of curriculum evalua ion. AERA Monograph Series on Curriculum Evaluation, No. L. Chicago: Rand-McNally, 1967. Pp. 13-18.

Tyler, R. W. (Ed.) Educational evaluation: New roles, new means. 68th Yearbook, Part II. Chicago: National Society for the Study of Education, 1969.

Wald, A. Sequential analysis. New York: Wiley, 1947.

Wilde, D. J., & Beightler, C. S. Foundations of optimization. Englewood Cliffs, N. J.: Prentice-Hall, 1967.

Wolf, M. M., & Risley, T. R. Reinforcement: Applied research. Paper presented at a Conference on the Nature of Reinforcement, University of Pittsburgh, 1969.

Woodrow, H. A. The effect of practice on groups of different initial ability. Journal of Educational Psychology, 1938, 29, 268-278.

Woodrow, H. A. The ability to learn. Psychological Review, 1946, 53, 147-158.

Zeaman, D., & Kaufman, H. Individual differences and theory in a motor learning task. Psychological Monographs, 1955, 69, (Whole No. 391).

Footnotes

1. The preparation of this chapter, which will appear in R. L. Thorndike (Ed.), Educational Measurement, was supported by the Personnel and Training Branch, Psychological Sciences Division, Office of Naval Research and by the Learning Research and Development Center supported as a research and development center by funds from the United States Office of Education, Department of Health, Education and Welfare.

2. The correlation coefficient has been the chief "measure" of the predictive validity of a test in the past. [Ed.: cross-reference to Cronbach's chapter if this point is discussed there.]

3. The term "mastery" means that an examinee makes a sufficient number of correct responses on the sample of test items presented to him in order to support the generalization (from this sample of items to the domain or universe of items implied by an instructional objective) that he has attained the desired, pre-specified degree of proficiency with respect to the domain. In certain situations, this can be considered as a simple or compound hypothesis testing situation.

4. Note that "item difficulty" has a meaning in this context only in references to the sequence or hierarchy which is employed. It is not used in the same way as in classical measurement theory (see Lord & Novick, 1968, pp. 328-329), although such uses coincide when a group of individuals who are heterogeneous with respect to the sequence are tested.

5. As indicated in Chapter 15, these decisions are terms placement decisions. The distinction between the use of these terms in that chapter and in this one have been pointed out (pp. 30-32).

TABLE 1

OBJECTIVES FOR COMPUTER-ASSISTED BRANCHED TESTING FOR ADDITION-SUBTRACTION

SUBTRACTION	BEHAVIOR
1	Solves subtraction problems related to single-digit combinations by multiples of ten.
2	Solves subtraction problems with no borrowing. Three- and four-digit combinations.
3	Solves subtraction problems from memory for two-digit sums less than or equal to twenty.
4	Subtracts two-digit numbers with borrowing from the tens' place.
5	Subtracts three-digit numbers with borrowing from the tens' <u>or</u> hundreds' place.
6	Subtracts three-digit numbers with borrowing from the tens' <u>and</u> hundreds' place.

ADDITION	BEHAVIOR
1	Solves addition problems from memory for sums less than or equal to twenty.
2	Solves subtraction problems from memory for sums less than or equal to nine.
3	Solves subtraction problems from memory for two-digit sums less than or equal to twenty.
4	Solves addition problems related to single-digit combinations by multiples of ten.
5	Finds the missing addend for problems with three single-digit addends.
6	Does column addition with no carrying. Two addends with three- and four-digit combinations.
7	Does column addition with no carrying. Three- or four-digit numbers with three to five addends.
8	Adds two-digit numbers with carrying to the tens' <u>or</u> hundreds' place. Two addends.
9	Finds the sums for column addition using three to five single-digit addends.
10	Adds two-digit numbers with carrying to the tens' <u>or</u> hundreds' place. Three or four addends.
11	Adds two-digit numbers with carrying to the tens' <u>and</u> hundreds' place. Two to four addends.
12	Adds three-digit numbers with carrying to the tens' <u>or</u> hundreds' place. Two to four addends.
13	Adds three-digit numbers with carrying to the tens' <u>and</u> hundreds' place. Two to four addends.

FIGURE CAPTIONS

Figure 1. Curriculum hierarchy on the addition of integers. (Reprinted from Gagné, Mayor, Garstens, and Paradise, 1962)

Figure 2. Curriculum hierarchy for counting a collection of movable objects. (Resnick, personal communication)

Figure 3. Curriculum hierarchy for placing an object in a two-dimensional matrix. (Resnick, personal communication)

Figure 4. Two possible hierarchies of sequence of instruction.

Figure 5. Illustration of alternative instructional sequences and some regression functions that may be useful in deciding a predictor test's value in making decisions concerning sequence allocation.

Figure 6. Hierarchies of objectives for an arithmetic unit in addition and subtraction. (Adapted from Ferguson, 1969)

Figure 7. Flow diagram for an instructional system. (Groen & Atkinson, 1966)

Figure 8. Instructional process flowchart for the IPI procedure. (Adapted from Lindvall, Cox, and Bolvin, 1970)

Figure 9. Examples of item forms from the subtraction universe. (Reprinted from Hively, Patterson, and Page, 1968)

Figure 10. Illustrated of Hively's task format and task generation rules. (From Hively, 1966b)

Figure 11. An example of a verbal replacement set for a variable element in an item form. (Adapted from Osburn, 1968)

Figure 12. Basic operational elements in development and evaluation of a system for IPI. (Adapted from Lindvall, Cox, and Bolvin, 1970)

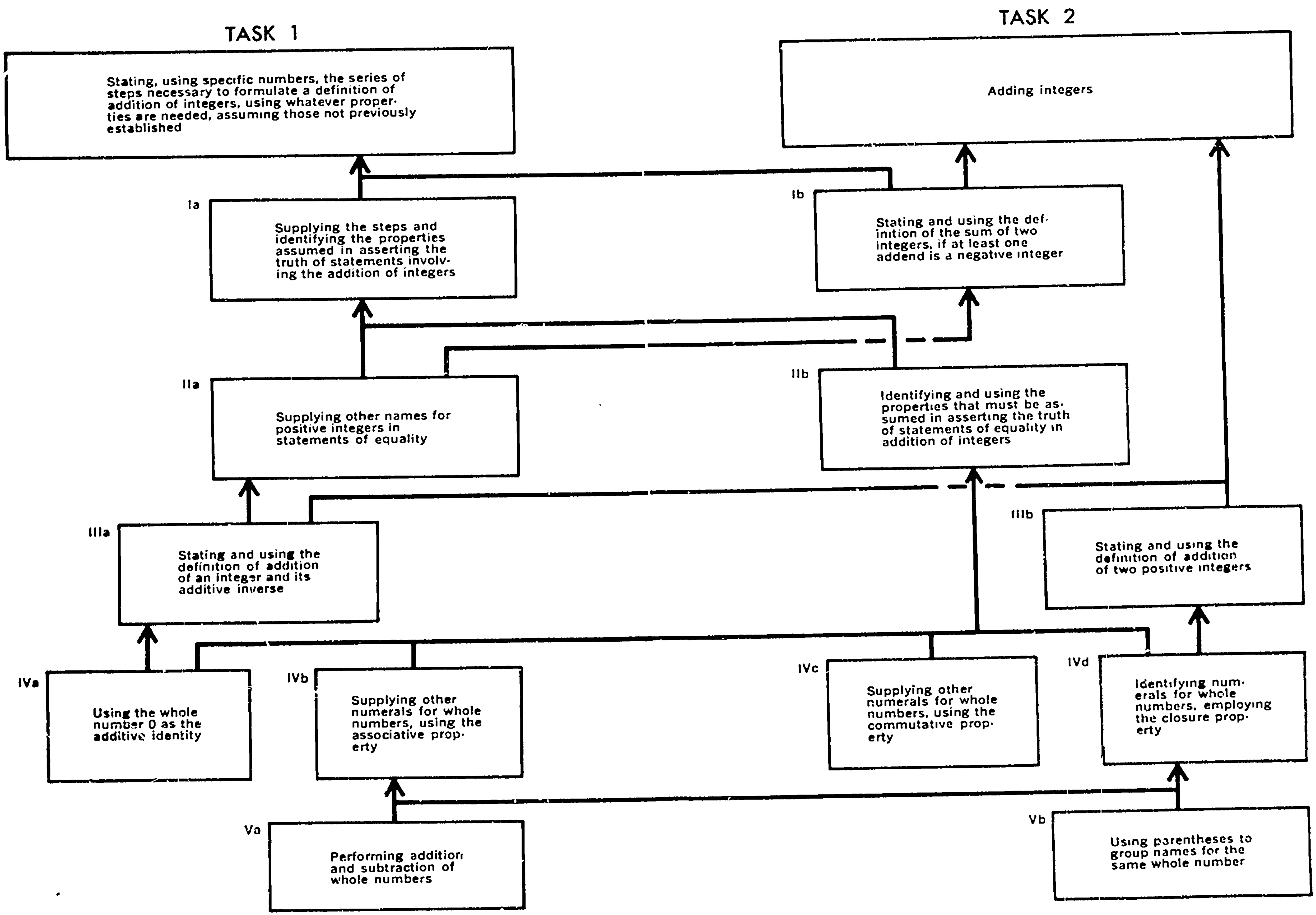

Figure 1. Curriculum hierarchy on the addition of integers.

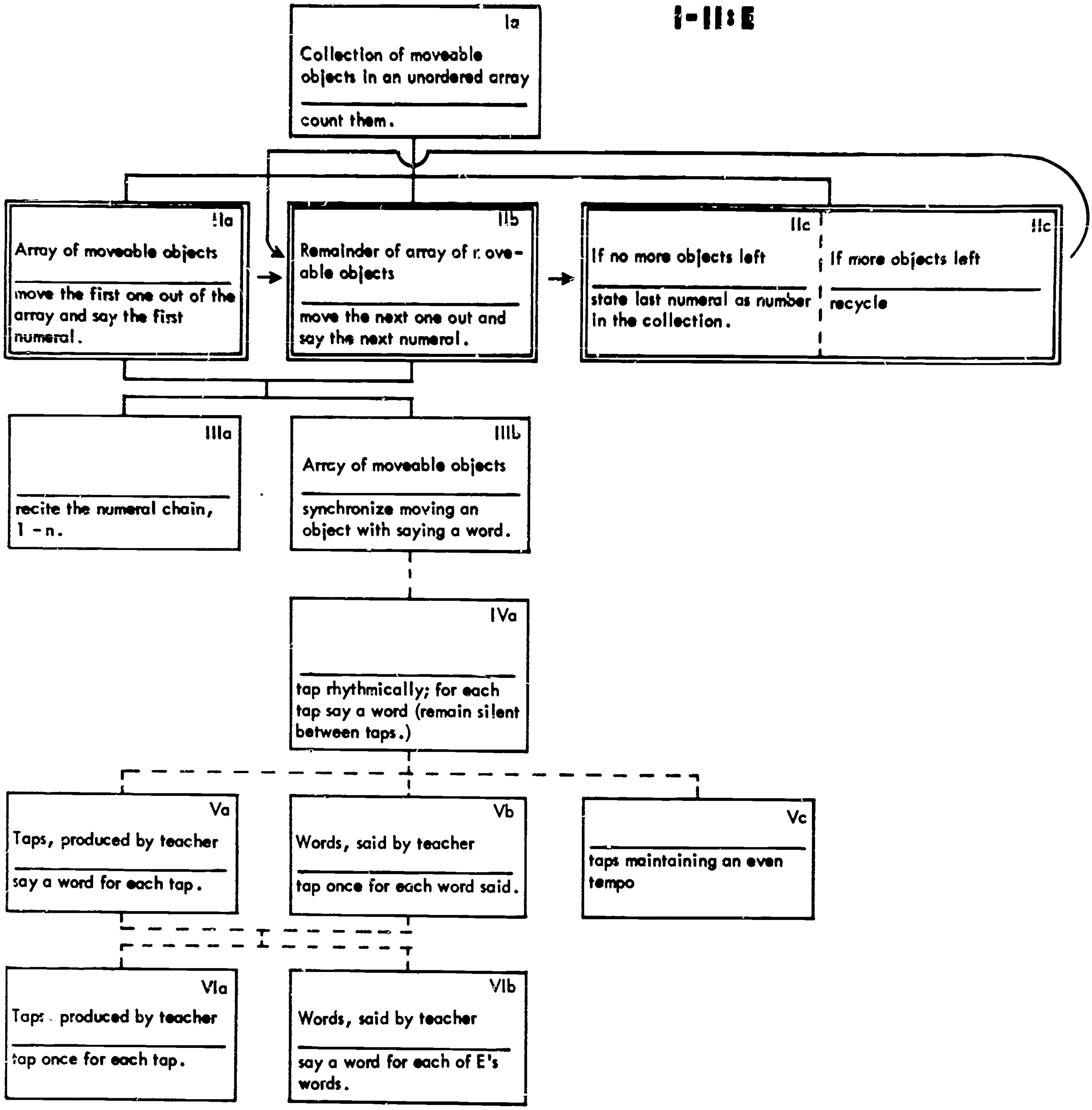

Figure 2. Curriculum hierarchy for counting a collection of movable objects.

CLASSIFICATION MATRIX

ANALYSIS 2

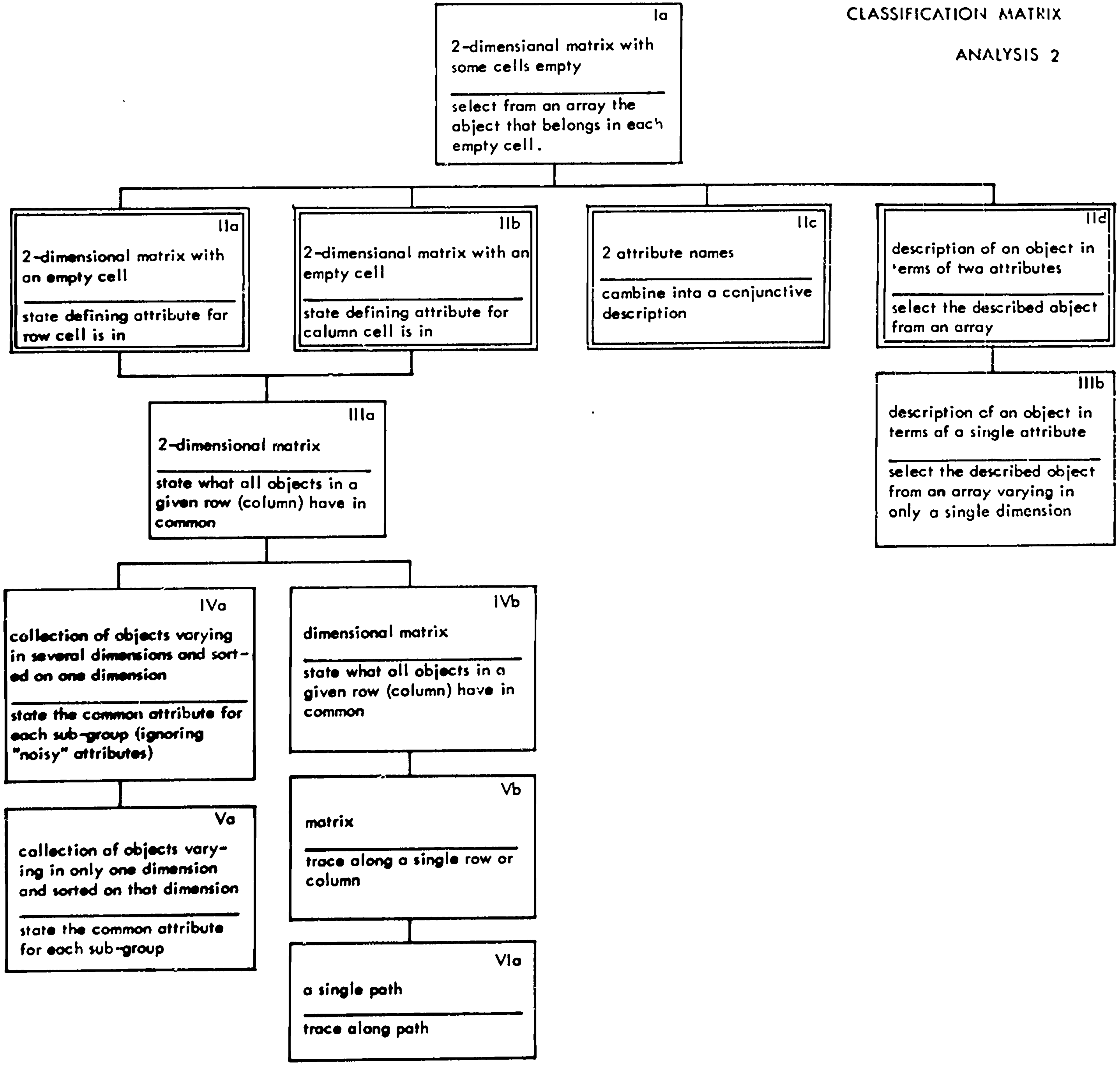

Figure 3. Curriculum hierarchy for placing an object in a two-dimensional matrix.

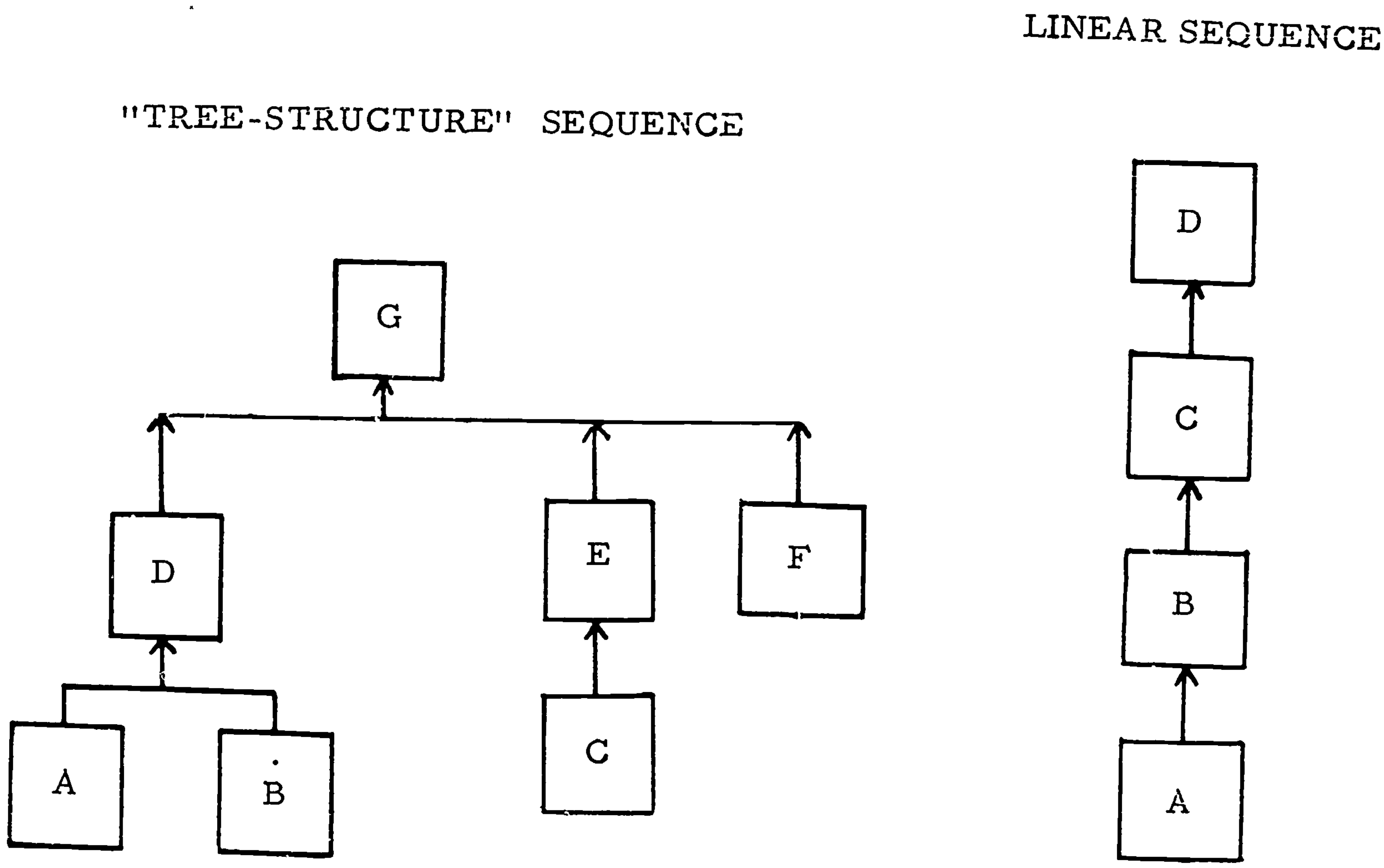

Figure 4. Two possible hierarchies of sequence of instruction.

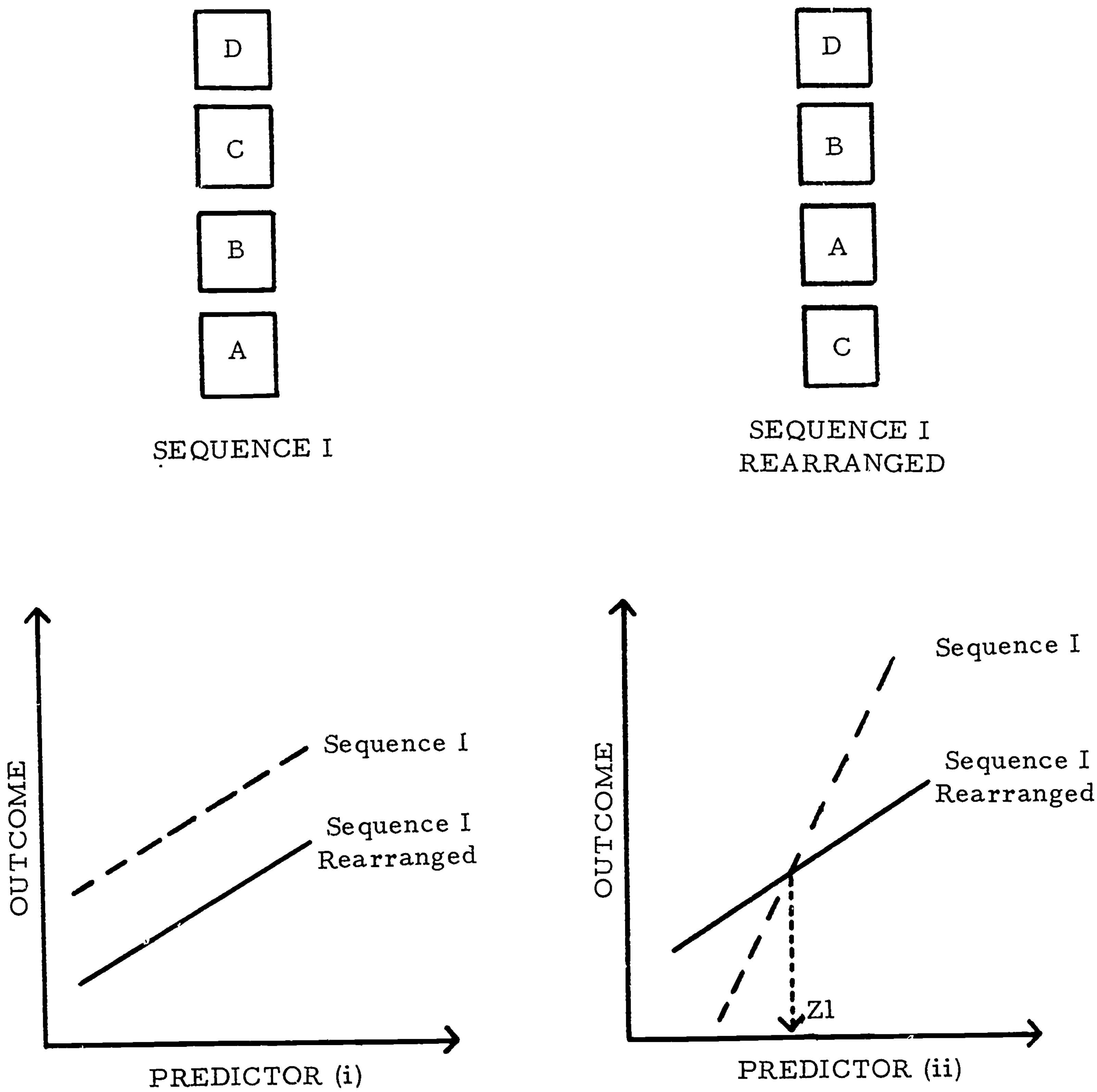

Figure 5. Illustration of alternative instructional sequences and some regression functions that may be useful in deciding a predictor test's value in making decisions concerning sequence allocation.

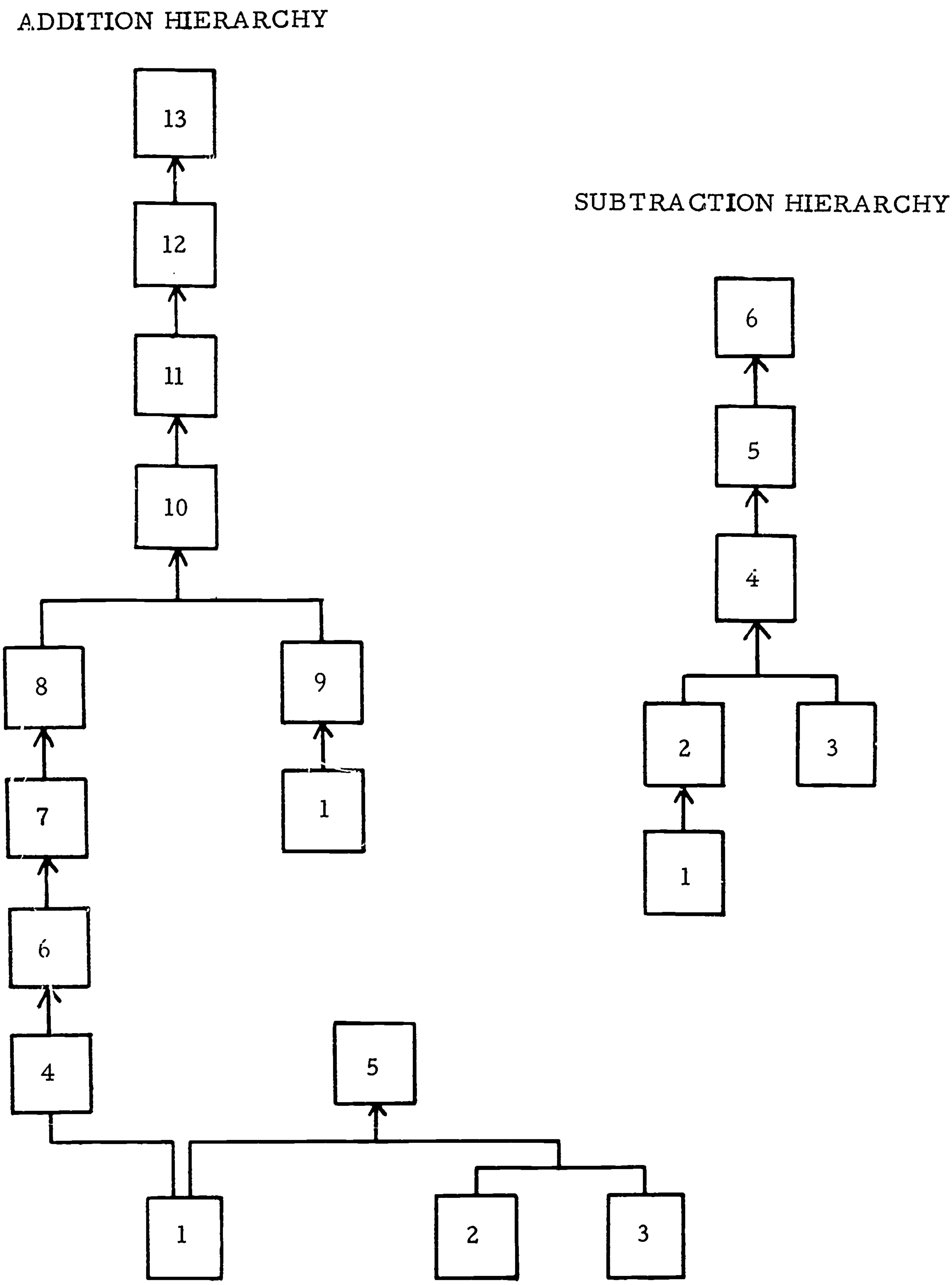

Figure 6. Hierarchies of objectives for an arithmetic unit in addition and subtraction.

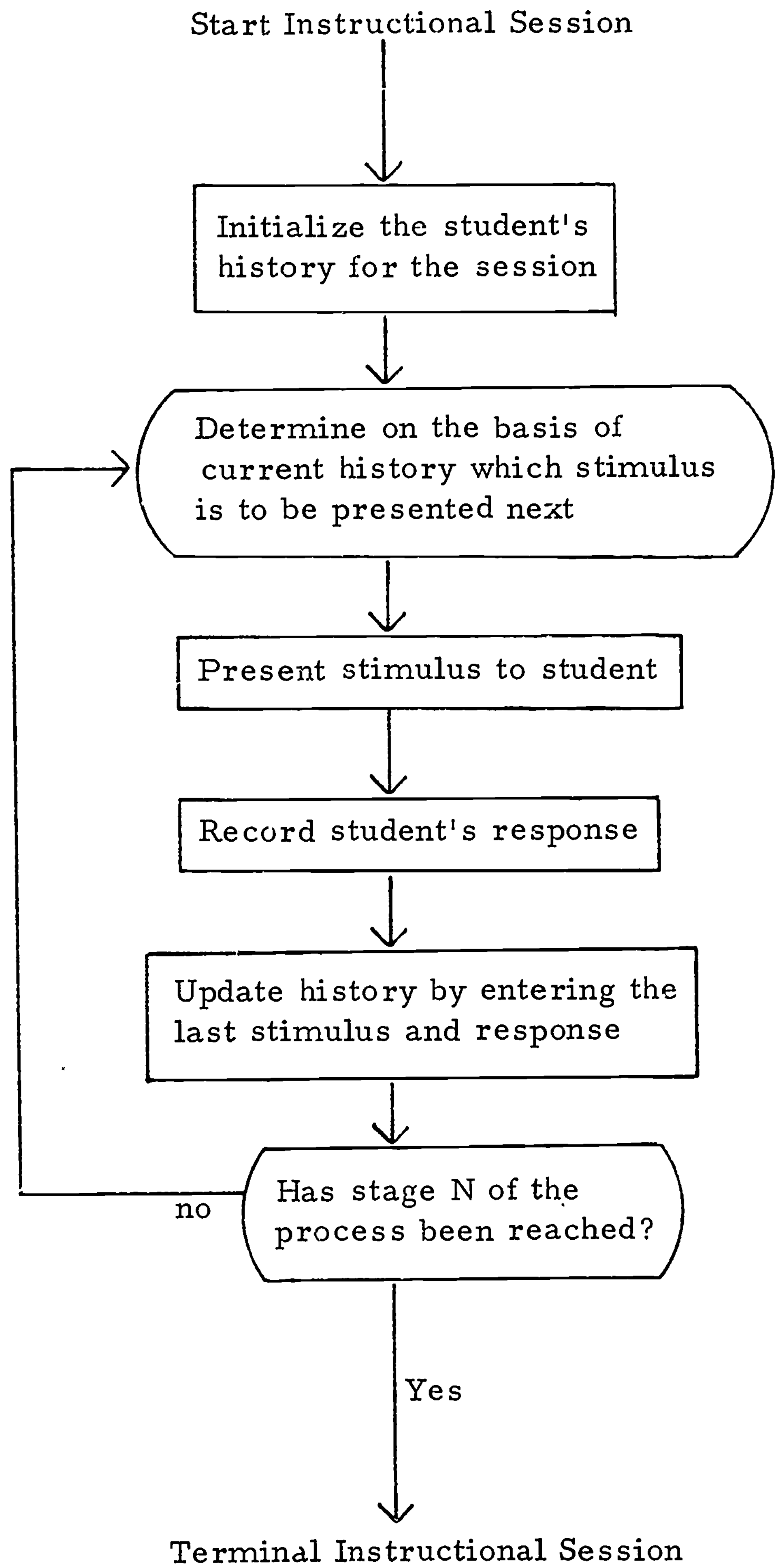

Figure 7. Flow diagram for an instructional system.

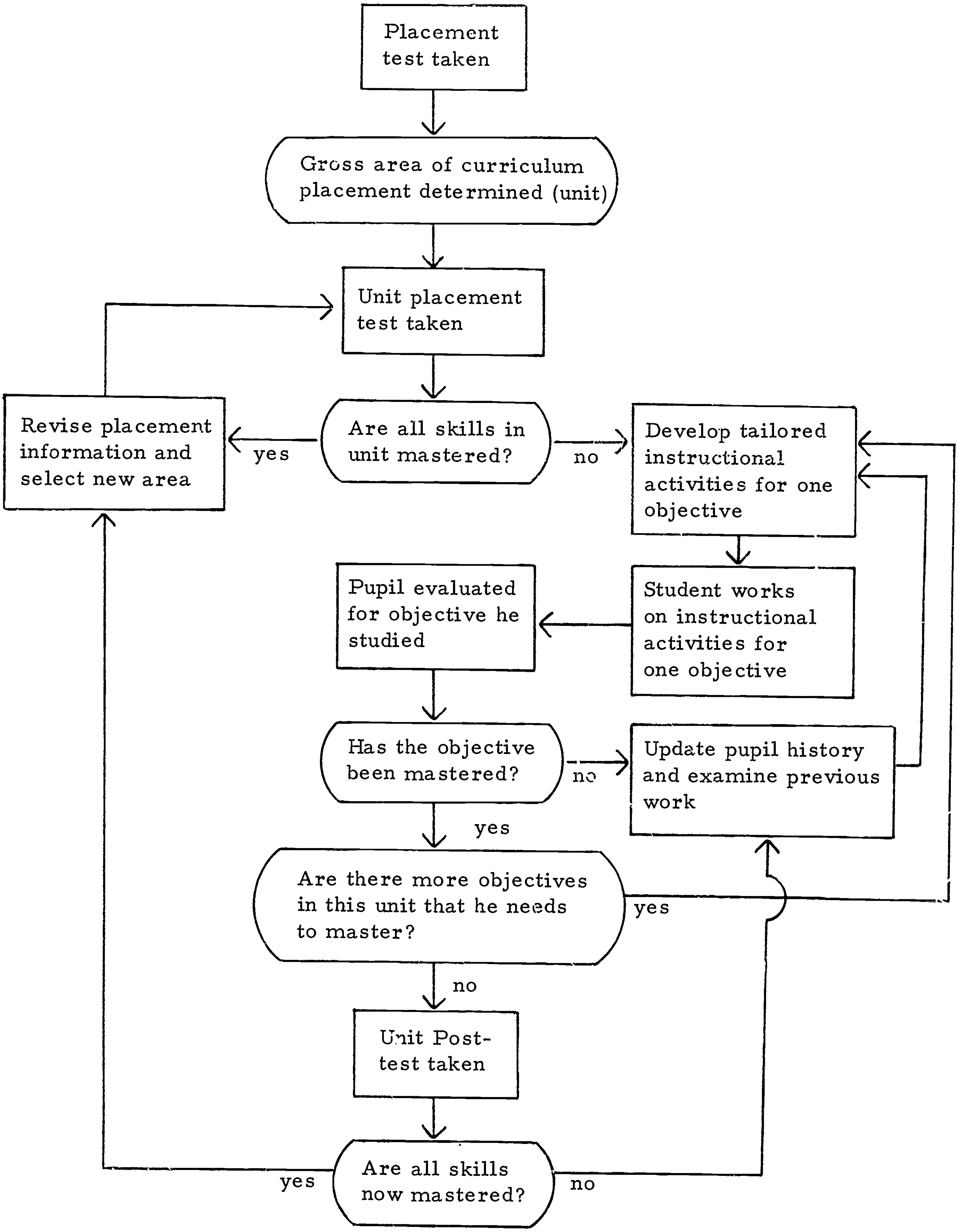

Figure 8. Instructional process flowchart for the IPI procedure.

Descriptive Title	Sample Item	General Form	Generation Rules[a]
Basic fact; minuend 10.	13 — 6	A —B	1. $A = 1a$; $B = b$ 2. $(a<b) \epsilon U$ 3. $\{H, V\}$
Simple borrow; one-digit subtrahend.	53 — 7	A —B	1. $A=a_1a_2$; $B=b$ 2. $a_1 \epsilon U - \{1\}$ 3. $(b>a_2) \epsilon U_0$
Borrow across 0.	403 —138	A —B	1. $N \epsilon \{3, 4\}$ 2. $A = a_1a_2 \ldots$; $B = b_1b_2 \ldots$ 3. $(a_1>b_1)$, $(a_3<b_3)$, $(a_4 \geq b_1) \epsilon U_0$ 4. $b_2 \epsilon U_0$ 5. $a_2=0$ 6. $P\{\{1, 2, 3\}, \{4\}\}$
Equation; missing subtrahend.	42—___=25	A—___=B	1. $A=a_1a_2$; $B=b_1b_2$ 2. $a_1 \epsilon U$ 3. $a_2, b_1, b_2 \epsilon U_0$ 4. *Check*: $O<B<A$

[a]Explanation of notation:

Capital letters A, B, . . . represent numerals.

Small letters (with or without subscripts) a. b. a_1. b_2. etc., represent digits.

$x \epsilon \{---\}$: Choose at random a replacement for x from the given set.

a, b, c, $\epsilon \{---\}$: All of a, b, c are chosen from the given set *with replacement*.

N_A: Number of digits in numeral A.

N: Number of digits in each numeral in the problem.

$a_1, a_2, \ldots \epsilon \{---\}$: Generate all the a necessary. In general " . . . " means continue the pattern established.

$(a< b) \epsilon \{---\}$: Choose two numbers at random *without* replacement; let *a* be the smaller.

$\{H, V\}$: Choose a horizontal or vertical format.

$P\{A, B. \ldots\}$: Choose a permutation of the elements in the set. (If the set consists of subscripts, permute those subscripted elements.)

Set operations are used as normally defined. Note that $A - B = A - \ddot{B}$. Ordered pairs are also used as usual.

Check: If a check is not fulfilled. regenerate all elements involved in the *check* statement (and any elements dependent upon them).

Special sets:

$U = \{1, 2, \ldots, 9\}$

$U_0 = \{0, 1. \ldots, 9\}$

Figure 9. Examples of item forms from the subtraction universe.

Purpose: To test the ability to solve an equality necessitating application of Theorem A,[1] Postulate B,[2] and Postulate C.[3] The solution set is to be non-empty and bounded by integers.

Task Format: $\underline{a} - \underline{b} \mid \underline{c} + (-1)^{\underline{f}}\underline{d} \mid \geq \underline{e}$

Generation Rules:

1. $\underline{c} \in \{x,y,z\}$
2. $\underline{b} \in \{2,3,4,5\}$
3. $\underline{f} \in \{0,1\}$
4. $\underline{d} \in \{1,2,3,\ldots,9\}$
5. $\underline{g} \in \{kb \mid k \in \{1,2,\ldots,5\} \text{ and } kb \neq b\}$
6. $\underline{e} \in \{1,2,3,\ldots,20\}$
7. $\underline{a} = \underline{g} + \underline{e}$

Explanation of Generation Rules

1. $\underline{c}$ is the variable of the inequality; x, y, or z may be used.

2. $\underline{b}$, the coefficient of the absolute value term, can vary from 2 to 5.

3. $(-1)^{f}$ allows the sign of the constant within the absolute value term to vary.

4. The constant $\underline{d}$ can vary from 1 to 9.

5. $\underline{g}$ is a multiple of $\underline{b}$, up to $5\underline{b}$, and not equal to $\underline{b}$.

6. $\underline{e}$ is any natural number from 1 to 20.

7. $\underline{a} = \underline{g} + \underline{e}$. In solving the problem, one will arrive at the step $\underline{a} - \underline{e} \geq \underline{b}|\underline{c} + (-1)^{f}\underline{d}|$. Since $\underline{a} - \underline{e} = \underline{g}$, and $\underline{g}$ is a multiple of $\underline{b}$, a cancellation step is required next. It is this pattern that must remain constant across the form.

---

[1] Theorem 1. If $\underline{a}$ is a real number and $\underline{a} > 0$, then $|x| < \underline{a}$ if and only if $-\underline{a}<x<\underline{a}$. [Use $\leq$, where x is of the form $(y \pm \underline{b})$.]

[2] Postulate 1. If $\underline{a}$, $\underline{b}$, $\underline{c}$ are real numbers such that $\underline{a} < \underline{b}$, then $\underline{a} + \underline{c} < \underline{b} + \underline{c}$. [Applied where $\underline{c}$ is a constant and also where $\underline{c}$ is the absolute value term.]

[3] Postulate 2. If $\underline{a}$, $\underline{b}$, $\underline{c}$ are real numbers such that $\underline{a}<\underline{b}$ and $0<\underline{c}$, then $\underline{ac}<\underline{bc}$

Figure 10. Illustrated of Hively's task format and task generation rules.

Item Form

Given ($ND:\mu, \sigma$) and (Region $ND:\mu, \sigma$). If one sample point (P) is randomly selected from ($ND:\mu, \sigma$), what is the probability that (P) is in (Region $ND:\mu, \sigma$)?

Possible Replacement Set for (ND: μ,σ).*

1. A fair penny is tossed (N) times and the number of heads is recorded.
2. John's true score on a certain test is (T) and the standard error of the test is (SE).
3. An urn contains (P) white balls and (Q) red balls. (N) balls are randomly selected with replacement and the number of white balls is noted.
4. The Wechsler Adult Intelligence Scale is standardized over the general population to mean of 100 and a standard deviation of 15.
5. A rat presses a bar an average of (P) times per minute when a light is on, and (Q) times per minute when the light is off. Under both conditions the distribution of bar presses is approximately normal with a standard deviation of (SD).
6. Sam takes a test consisting of (R) (K)-alternative multiple choice items and guesses on all items.
7. A certain batch of ball bearings is known to contain 20 per cent defectives. (N) ball bearings are shipped to a customer.
8. A certain test contains (R) items that are all of equal difficulty, $P = (X)$, for a population of 9th grade students.
9. A white die is rolled (N) times and the number of times the (Y)-face turns up is noted.
10. A certain firm produces packaged butter. Quality control has shown that the average weight per package is 16.5 ounces with a standard deviation of .5 ounces.

*Before this item form can be used to generate items, suitable numerical replacement sets need to be defined.

Figure 11. An example of a verbal replacement set for a variable element in an item form.

INSTRUCTIONAL OBJECTIVES that:
(a) can be used by lesson writers, test developers, and teachers without ambiguity.
(b) are in prerequisite order as evidenced by pupil mastery and progression.
(c) permit lesson writers to develop sequences of lessons that have no missing steps nor overlapping steps and with which pupils can make progress.
(d) are such that persons can agree as to what the pupil is to be taught and on what he is to be tested.
(e) are inclusive enough so that no important gaps in abilities taught are discovered.

THE TESTING PROGRAM:
(a) is used to place pupils at correct points in the instructional continua.
(b) provides valid diagnosis of pupil needs.
(c) provides a valid assessment of mastery of objectives and of units.
(d) is administered so that the pupil is taking CET's and unit tests at proper times.
(e) provides data that are found useful by the teachers for developing valid prescriptions.
(f) provides data that are meaningful to the student.

INSTRUCTIONAL PRESCRIPTIONS:
(a) are based upon proper use of test results and specified prescription writing procedures.
(b) provide learning experiences that are a challenge but permit regular progress.
(c) vary from pupil to pupil depending upon individual differences.
(d) permit pupil to proceed at his best rate.
(e) are interpreted and used correctly by the pupil.
(f) are modified as required.

THE INSTRUCTIONAL MATERIALS AND DEVICES:
(a) are easily identified with the proper objective.
(b) have demonstrated instructional effectiveness.
(c) are used by pupils largely in individual independent study.
(d) are used by pupils in individualized packages.
(e) keep the pupil actively involved.
(f) require a minimum of direct teacher help to pupils.
(g) are shown to teach more effectively as they are revised.

THE TEACHER CLASSROOM ACTIVITIES are such that:
(a) there is little delay in the pupil's getting help when he needs it.
(b) teacher assistance to pupils is largely on an individual basis.
(c) the teacher will spend some class time in examining pupil work and in developing prescriptions.
(d) positive reinforcement of desirable behavior is employed.
(e) teachers give the students considerable freedom.
(f) little time is spent on lectures (etc.) to the group, and individual or small group tutoring is employed.

PUPIL CLASSROOM ACTIVITIES are such that:
(a) pupils work largely on an individual and independent basis.
(b) pupils are studying with a minimum of wasted time.
(c) pupils secure needed materials in an efficient manner.
(d) pupils help each other on occasion.

CLASSROOM MANAGEMENT PROCEDURES are such that:
(a) teacher aides score papers and record results in an efficient manner.
(b) pupils score some work pages.
(c) pupils procure own lesson materials.
(d) pupils decide when to have lessons scored.

Figure 12. Basic operational elements in development and evaluation of a system for IPI.

ONR Distribution List

NAVY

4 Chief of Naval Research
Code 458
Department of the Navy
Washington, D. C. 20360

1 Director
ONR Branch Office
495 Summer Street
Boston, Massachusetts 02210

1 Director
ONR Branch Office
219 South Dearborn Street
Chicago, Illinois 60604

1 Director
ONR Branch Office
1030 East Green Street
Pasadena, California 91101

6 Director
Naval Research Laboratory
Washington, D. C. 20390
Attn: Technical Information Division

6 Director
Naval Research Laboratory
Attn: Library, Code 2029 (ONRL)
Washington, D. C. 20390

1 Office of Naval Research
Area Office
207 West Summer Street
New York, New York 10011

1 Office of Naval Research
Area Office
1076 Mission Street
San Francisco, California 94103

20 Defense Documentation Center
Cameron Station, Building 5
5010 Duke Street
Alexandria, Virginia 22314

1 Superintendent
Naval Postgraduate School
Monterey, California 93940
Attn: Code 2124

1 Head, Psychology Branch
Neuropsychiatric Service
U. S. Naval Hospital
Oakland, California 94627

1 Commanding Officer
Service School Command
U. S. Naval Training Center
San Diego, California 92133

3 Commanding Officer
Naval Personnel & Training
Research Laboratory
San Diego, California 92152

1 Officer in Charge
Naval Medical Neuropsychiatric
Research Unit
San Diego, California 92152

NAVY

1 Commanding Officer
Naval Air Technical Training Center
Jacksonville, Florida 32213

1 Dr. James J. Regan
Naval Training Device Center
Orlando, Florida 32813

1 Chief
Aviation Psychology Division
Naval Aerospace Medical Institute
Naval Aerospace Medical Center
Pensacola, Florida 32512

1 Chief
Naval Air Reserve Training
Naval Air Station
Box 1
Glenview, Illinois 60026

1 Technical Library
U. S. Naval Weapons Laboratory
Dahlgren, Virginia 22448

1 Chairman
Leadership/Management Committee
Naval Sciences Department
U. S. Naval Academy
Annapolis, Maryland 21402

1 Dr. A. L. Slafkosky
Scientific Advisor
Commandant of the Marine Corps
(Code AX)
Washington, D. C. 20380

1 Technical Services Division
National Library of Medicine
8600 Rockville Pike
Bethesda, Maryland 20014

1 Behavioral Sciences Department
Naval Medical Research Institute
National Naval Medical Center
Bethesda, Maryland 20014

1 Commanding Officer
Naval Medical Field Research Laboratory
Camp Lejeune, North Carolina 28542

1 Director
Aerospace Crew Equipment Department
Naval Air Development Center, Johnsville
Warminster, Pennsylvania 18974

1 Chief
Naval Air Technical Training
Naval Air Station
Memphis, Tennessee 38115

1 Technical Library
Naval Training Device Center
Orlando, Florida 32813

1 Technical Library
Naval Ship Systems Command
Main Navy Building, Rm. 1532
Washington, D. C. 20360

1 Technical Library
Naval Ordnance Station
Indian Head, Maryland 20640

1 Naval Ship Engineering Center
Philadelphia Division
Technical Library
Philadelphia, Pennsylvania 19112

1 Library, Code 0212
Naval Postgraduate School
Monterey, California 93940

NAVY

1 Commander
Operational Test &
Evaluation Force
U. S. Naval Base
Norfolk, Virginia 23511

1 Office of Civilian
Manpower Management
Department of the Navy
Washington, D. C. 20350
Attn: Code 023

1 Chief of Naval Operations, Op-07TL
Department of the Navy
Washington, D. C. 20350

1 Chief of Naval Material
(MAT 031M)
Room 1323, Main Navy Bldg.
Washington, D. C. 20360

1 Naval Ship Systems Command
Code 03H
Department of the Navy
Main Navy Building
Washington, D. C. 20360

1 Chief
Bureau of Medicine and Surgery
Code 513
Washington, D. C. 20390

1 Technical Library
Bureau of Naval Personnel
(Pers-11b)
Department of the Navy
Washington, D. C. 20370

1 Director
Personnel Research &
Development Laboratory
Washington Navy Yard, Building 200
Washington, D. C. 20390

1 Commander, Naval Air Systems Command
Navy Department, AIR-4133
Washington, D. C. 20360

1 Commandant of the Marine Corps
Headquarters, U. S. Marine Corps
Code A01B
Washington, D. C. 20380

ARMY

1 Human Resources Research Office
Division #6, Aviation
Post Office Box 428
Fort Rucker, Alabama 36360

1 Human Resources Research Office
Division #3, Recruit Training
Post Office Box 5787
Presidio of Monterey, California 93940
Attn: Library

1 Human Resources Research Office
Division #4, Infantry
Post Office Box 2086
Fort Benning, Georgia 31905

1 Department of the Army
U. S. Army Adjutant General School
Fort Benjamin Harrison, Ind. 46216
Attn: AGCS-EA

NAVY

1 Technical Reference Library
Naval Medical Research Institute
National Naval Medical Center
Bethesda, Maryland 20014

1 Technical Library
Naval Ordnance Station
Louisville, Kentucky 40214

1 Library
Naval Electronics
Laboratory Center
San Diego, California 92152

1 Technical Library
Naval Undersea Warfare Center
3202 E. Foothill Boulevard
Pasadena, California 91107

1 AFHRL (HRTT/Dr. Ross L. Morgan)
Wright-Patterson Air Force Base
Ohio 45433

1 AFHRL (HRO/Dr. Meyer)
Brooks Air Force Base
Texas 78235

1 Mr. Michael Macdonald-Ross
Instructional Systems Associates
West One
49 Welbeck Street
London W1M 7HE
England

1 Commanding Officer
U. S. Naval Schools Command
Mare Island
Vallejo, California 94592

1 Dr. Don C. Coombs
Assistant Director
ERIC Clearinghouse
Stanford University
Palo Alto, California 94305

1 Scientific Advisory Team (Code 71)
Staff, COMASWFORLANT
Norfolk, Virginia 23511

1 ERIC Clearinghouse
Educational Media and Technology
Stanford University
Stanford, California

1 ERIC Clearinghouse
Vocational and Technical Education
Ohio State University
Columbus, Ohio 43212

1 Education & Training Developments Staff
Personnel Research & Development Lab.
Building 200, Washington Navy Yard
Washington, D. C. 20390

1 Director
Education & Training Sciences Dept.
Naval Medical Research Institute
Building 142
National Naval Medical Center
Bethesda, Maryland 20014

1 LCDR J. C. Meredith, USM (Ret.)
Institute of Library Research
University of California, Berkeley
Berkeley, California 94720

1 Mr. Joseph B. Blankenheim
NAVELEX 0474
Munitions Building, Rm. 3721
Washington, D. C. 20360

ARMY

1 Director of Research
U. S. Army Armor
Human Research Unit
Fort Knox, Kentucky 40121
Attn: Library

1 Research Analysis Corporation
McLean, Virginia 22101
Attn: Library

1 Human Resources Research Office
Division #5, Air Defense
Post Office Box 6021
Fort Bliss, Texas 79916

1 Human Resources Research Office
Division #1, Systems Operations
300 North Washington Street
Alexandria, Virginia 22314

1 Director
Human Resources Research Office
The George Washington University
300 North Washington Street
Alexandria, Virginia 22314

1 Armed Forces Staff College
Norfolk, Virginia 23511
Attn: Library

1 Chief
Training and Development Division
Office of Civilian Personnel
Department of the Army
Washington, D. C. 20310

1 U. S. Army Behavioral Science
Research Laboratory
Washington, D. C. 20315

1 Walter Reed Army Institute of Research
Walter Reed Army Medical Center
Washington, D. C. 20012

1 Behavioral Sciences Division
Office of Chief of Research
and Development
Department of the Army
Washington, D. C. 20310

1 Dr. George S. Harker
Director, Experimental Psychology Div.
U. S. Army Medical Research Laboratory
Fort Knox, Kentucky 40121

AIR FORCE

1 Director
Air University Library
Maxwell Air Force Base
Alabama 36112
Attn: AUL-8110

1 Cadet Registrar
U. S. Air Force Academy
Colorado 80840

1 Headquarters, ESD
ESVPT
L. G. Hanscom Field
Bedford, Massachusetts 01731
Attn: Dr. Mayer

1 AFHRL (HRT/Dr. G. A. Eckstrand)
Wright-Patterson Air Force Base
Ohio 45433

AIR FORCE

1 Commandant
U. S. Air Force School of
Aerospace Medicine
Brooks Air Force Base, Texas 78235
Attn: Aeromedical Library
(SMSDL)

1 6570th Personnel Research Laboratory
Aerospace Medical Division
Lackland Air Force Base
San Antonio, Texas 78236

1 AFOSR (SRLB)
1400 Wilson Boulevard
Arlington, Virginia 22209

1 Research Psychologist
SCBB, Headquarters
Air Force Systems Command
Andrews Air Force Base
Washington, D. C. 20331

1 Headquarters, U. S. Air Force
Chief, Analysis Division (AFPDPL)
Washington, D. C. 20330

1 Headquarters, U. S. Air Force
Washington, D. C. 20330
Attn: AFPTRTB

1 Headquarters, U. S. Air Force
AFRDDG
Room 1D373, The Pentagon
Washington, D. C. 20330

1 Headquarters, USAF (AFPTRD)
Training Devices and Instructional
Technology Division
Washington, D. C. 20330

MISCELLANEOUS

1 Dr. Alvin E. Goins, Executive Secretary
Personality & Cognition Research
Review Committee
Behavioral Sciences Research Branch
National Institute of Mental Health
5454 Wisconsin Avenue, Room 10A11
Chevy Chase, Maryland 20203

1 Dr. Mats Bjorkman
University of Umea
Department of Psychology
Umea 6, Sweden

1 Technical Information Exchange
Center for Computer Sciences
and Technology
National Bureau of Standards
Washington, D. C. 20234

1 Director
Defense Atomic Support Agency
Washington, D. C. 20305
Attn: Technical Library

1 Executive Secretariat
Interagency Committee on
Manpower Research
Room 515
1738 "M" Street, N. W.
Washington, D. C. 20036
(Attn: Mrs. Ruth Relyea)

1 Mr. Joseph J. Cowan
Chief, Personnel Research Branch
U. S. Coast Guard Headquarters
PO-1, Station 3-12
1300 "E" Street, N. W.
Washington, D. C. 20226

MISCELLANEOUS

1 Dr. Henry S. Odbert
National Science Foundation
1800 "G" Street, N. W.
Washington, D. C. 20550

1 Dr. Gabriel D. Ofiesh
Center for Educational Technology
Catholic University
4001 Harewood Road, N. E.
Washington, D. C. 20017

1 Dr. Joseph W. Rigney
Electronics Personnel Research Group
University of Southern California
University Park
Los Angeles, California 90007

1 Dr. Arthur I. Siegel
Applied Psychological Services
Science Center
404 East Lancaster Avenue
Wayne, Pennsylvania 19087

1 Dr. Arthur W. Staats
Department of Psychology
University of Hawaii
Honolulu, Hawaii 96822

1 Dr. Lawrence M. Stolurow
Harvard Computing Center
6 Appian Way
Cambridge, Massachusetts 02138

1 Dr. Ledyard R. Tucker
Department of Psychology
University of Illinois
Urbana, Illinois 61801

1 Dr. Benton J. Underwood
Department of Psychology
Northwestern University
Evanston, Illinois 60201

1 Dr. Joseph A. Van Campen
Institute for Math Studies in the
Social Sciences
Stanford University
Stanford, California 94305

1 Dr. John Annett
Department of Psychology
Hull University
Yorkshire
England

1 Dr. M. C. Shelesnyak
Interdisciplinary Communications Program
Smithsonian Institution
1025 Fifteenth Street, N. W.
Suite 700
Washington, D. C. 20005

1 Dr. Lee J. Cronbach
School of Education
Stanford University
Stanford, California 94305

1 Dr. John C. Flanagan
Applied Institutes for Research
P. O. Box 1113
Palo Alto, California 94302

1 Dr. M. D. Havron
Human Sciences Research, Inc.
Westgate Industrial Park
7710 Old Springhouse Road
McLean, Virginia 22101

1 Dr. Roger A. Kaufman
Department of Education
Institute of Instructional System
Technology & Research
Chapman College
Orange, California 92666

MISCELLANEOUS

1 Executive Officer
American Psychological Association
1200 Seventeenth Street, N. W.
Washington, D. C. 20036

1 Mr. Edmund C. Berkeley
Information International, Inc.
545 Technology Square
Cambridge, Massachusetts 02139

1 Dr. Donald L. Bitzer
Computer-Based Education Research Laboratory
University of Illinois
Urbana, Illinois 61801

1 Dr. C. Victor Bunderson
Computer Assisted Instruction Lab.
University of Texas
Austin, Texas 78712

1 Dr. F. J. DiVesta
Education & Psychology Center
Pennsylvania State University
University Park, Pennsylvania 16802

1 Dr. Phillip H. DuBois
Department of Psychology
Washington University
Lindell & Skinker Boulevards
St. Louis, Missouri 63130

1 Dr. Wallace Feurzeig
Bolt, Beranek & Newman, Inc.
50 Moulton Street
Cambridge, Massachusetts 02138

1 Dr. Carl E. Helm
Department of Educational Psychology
Graduate Center
City University of New York
33 West 42nd Street
New York, New York 10036

1 Dr. Bert Green
Department of Psychology
John Hopkins University
Baltimore, Maryland 21218

1 Dr. J. P. Guilford
University of Southern California
3551 University Avenue
Los Angeles, California 90007

1 Dr. Harold Gulliksen
Department of Psychology
Princeton University
Princeton, New Jersey 08540

1 Dr. Duncan N. Hansen
Center for Computer Assisted Instruction
Florida State University
Tallahassee, Florida 32306

1 Dr. Albert E. Hickey
Entelek, Incorporated
42 Pleasant Street
Newburyport, Massachusetts 01950

1 Dr. Howard H. Kendler
Department of Psychology
University of California
Santa Barbara, California 93106

1 Dr. Robert R. Mackie
Human Factors Research, Inc.
6780 Cortona Drive
Santa Barbara Research Park
Goleta, California 93107

1 S. Fisher, Research Associate
Computer Facility
Graduate Center
City University of New York
33 West 42nd Street
New York, New York 10036

Unclassified
Security Classification

DOCUMENT CONTROL DATA - R & D

Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Learning Research and Development Center University of Pittsburgh Pittsburgh, Pennsylvania 15213	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP

3. REPORT TITLE

Measurement in Learning and Instruction

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)
Technical Report

5. AUTHOR(S) (First name, middle initial, last name)

Robert Glaser and Anthony J. Nitko

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
March 1970	115	119

8a. CONTRACT OR GRANT NO. Nonr-624(18) b. PROJECT NO c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) None 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report) None

10. DISTRIBUTION STATEMENT
This document has been approved for public release and sale; its distribution is unlimited. Reproduction in whole or in part is permitted for any purpose of the U. S. Government.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	Personnel and Training Branch Psychological Sciences Division Office of Naval Research

13. ABSTRACT

The authors initially consider three general classes of instruction models found in current educational practice. One particular model of instruction--a general model for individualization and adapting instruction to individual differences--is described, and its testing and measurement implications are discussed. Central to this approach is the specification of desired instructional goals in terms of organizable domains of human performance criteria as well as adaptation of instruction on an individual basis so that these desired goals are attained by a maximum number of students. The description of the instructional model is followed by considerations relevant to the analysis of performance domains, individual assignment to instructional alternatives, and necessity for measuring what is learned by means of criterion-referenced tests. The last section of this chapter briefly discusses the important topic of evaluating and improving an instructional system and its components.

Unclassified
Security Classification

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Educational Measurement						
Individualization						
Criterion-Referenced Measurement						
Formative evaluation						
Computer-Assisted Testing						
Component Task Analysis						

DD FORM 1 NOV 65 1473 (BACK)
S/N 0101-807-6821

Unclassified
Security Classification